Network Diversity, Market Entry, and the Global Internet Backbone

El Hadi Caoui — U of Toronto

Andrew Steck — U of Toronto

November 3, 2023*

Abstract

In supply networks, path diversity is an important determinant of demand, entry, and firms' market power. We investigate this role in the context of the internet backbone, the worldwide network of undersea fiber-optic cables that underpins the internet. We specify a model of inter-regional bandwidth demand and cable operators' dynamic entry and supply choices. The model is estimated using novel data on cross-border data flows, prices, and cable characteristics. Counterfactual analysis reveals that preferences for path diversity account for a large share of entry rates and surplus created between 2005 and 2021. Relative to the socially optimal level of entry, distortions due to entrants' inability to capture fully the social benefits of diversity are as large as distortions due to business stealing.

Keywords: diversification, supply disruptions, internet backbone, entry, dynamic games

JEL Classification: L13, L96, L86

*Caoui: elhadi.caoui@rotman.utoronto.ca. Steck: andrew.steck@rotman.utoronto.ca. We thank Ambarish Chandra, Avi Goldfarb, Shane Greenstein, Matthew Mitchell, Andrew Sweeting, Joel Waldfogel, Xiaosong Wu, Biliana Yontcheva, seminar and conference participants at the University of Toronto, University of Maryland, NYU Stern, Cornell Johnson, IIOC, WEAI, CEA, CEPR Applied IO Conference, BSE Summer Forum, CRESSE, and NYU Stern Economics of Strategy workshop for helpful comments and discussions. We are grateful to Lilian Lea and Alan Mauldin from TeleGeography for helping us access the data and answering our questions about the global undersea internet cable industry. The analysis, findings, and conclusions expressed in this paper should not be attributed to TeleGeography.

1 Introduction

Supply disruptions are a prevalent concern in many industries, where natural disasters or production failures can have economically devastating consequences. Recent examples highlighting the vulnerability of supply networks abound. In March 2021, the container ship *Ever Given* ran aground and blocked the Suez Canal for over a week; a political protest in February 2022 shut down traffic on the Ambassador Bridge, which carries 25% of trade between the U.S. and Canada; in January 2022, the Hunga Tonga–Hunga Ha'apai volcanic eruption severed internet links to Tonga for over a month; and disruptions due to COVID-19 mitigation measures and outbreaks significantly disrupted chip supply in industries from graphic cards to automobiles.[1] In light of these disruptions, a natural response is supplier diversification, be it through sourcing from multiple suppliers or implementing alternative network links.[2] In this study, we examine empirically how diversification affects firms' incentives to enter and compete, and more broadly, its long-run impact on market structure. We do so in an important setting: the undersea telecommunications cables that comprise the global "backbone of the internet."

The effect of supplier diversification on market structure operates through two primary channels. First, as long as the disruptions of different suppliers are not perfectly correlated, diversification increases the degree of horizontal differentiation between suppliers; this differentiation insulates suppliers from competitive pressure. Second, for a fixed level of prices, market entry by additional suppliers will increase aggregate quantities demanded, a "market expansion" effect; intuitively, the value of the "portfolio" of the market supply increases with the number of competitors.

Our empirical application focuses on the global internet backbone from 2005 to 2021. This worldwide network of undersea fiber-optic cables comprises the primary means of intercontinental information transport, carrying more than 98% of all international internet traffic (data, video calls, instant messages and emails). Operational undersea cable networks are essential to well-functioning global economic and financial systems: e.g., the U.S.

[1] On February 24, 2021, U.S. President Biden signed Executive Order 14017, titled "America's Supply Chains," which directed the U.S. government to assess key U.S. supply chains to identify potential risks and weaknesses and create a plan to enhance resilience. Among the facets of this industrial policy initiative are $76 billion in tax credits and subsidies to boost domestic chip production and diminish reliance on East Asian suppliers. Other recent examples of supply disruptions include the sabotage and subsequent underwater gas leaks that occurred on the Nord Stream 1 and Nord Stream 2 natural gas pipelines in September 2022.

[2] The term "supplier diversification" refers to the practice of spreading the sourcing of goods and services across multiple suppliers, in order to reduce the risk of supply disruptions caused by the failure or disruption of a single supplier. The term "network diversity" refers to the characteristic of a supply network which implements multiple independent paths for goods or data to flow. In this sense, network diversity is one strategy to enhance a given network's resilience.

Clearing House Interbank Payment System processes more than $10 trillion per day in transactions with more than 22 economies via undersea cables (Federal Communications Commission (2015)). Undersea cables also feature prominently in intra-continental and even intra-national communication networks. Figure 1 shows the global undersea cable network in 2022.[3]

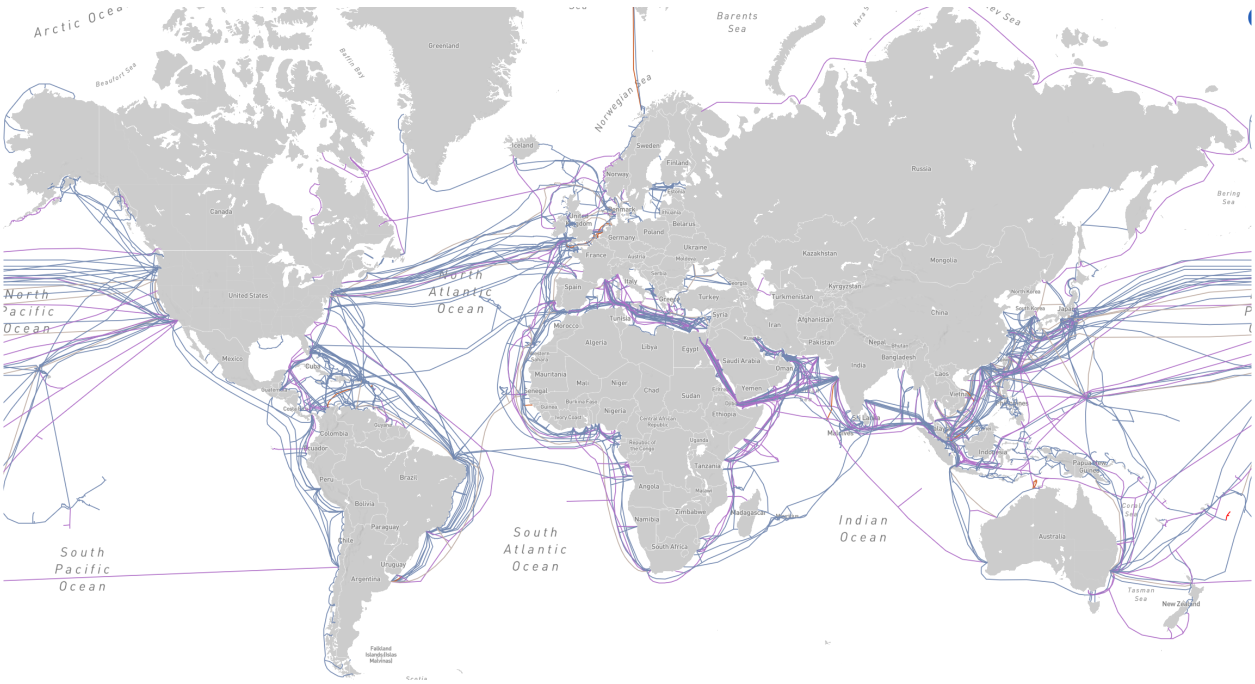


Figure 1: The Global Internet Backbone (source: Infrapedia.com)

The undersea internet cable industry is well-suited to examine the effect of the demand for diversification on market structure. Hundreds of cable failures and repairs take place every year, and the industry therefore places a high value on network resilience and path diversity. Industry estimates put the financial impact from interruptions of undersea fiber-optic systems in excess of $1.5 million per hour (Malphrus (2009)).[4] In addition, the industry has witnessed significant growth over the past twenty years, providing many instances of entry in a variety of market conditions. Finally, the undersea cable industry is of interest in its own right as a critical component of the "industrial organization of the internet supply chain."

[3]We do not consider the network of terrestrial cables, arguably another important component of the internet backbone, due to a lack of data on terrestrial cables and the relatively more complex topology of the terrestrial cable network.

[4]High-profile examples of cable outages include a powerful earthquake off Southern Taiwan which cut 9 cables on December 26, 2006 and took 11 repair ships 49 days to restore. The earthquake affected Internet links, financial markets, banking, airline bookings and general communications in China, Hong Kong, India, Singapore, Taiwan, Japan and the Philippines; the AAE-1 cable cut in June 2022 which caused widespread outages in Europe, East Africa, the Middle East, and South Asia and affected various Cloud service providers.

We leverage comprehensive data on this industry, which were obtained from TeleGeography, a telecommunications market research firm. The data provide detailed information on undersea cable systems (e.g., construction costs, cable length, ready-for-service dates, landing stations, capacities), as well as extensive quarterly panels on bandwidth prices (at the city-pair level) and data flows (at the country-pair level). To the authors' knowledge, these data have not previously been used in the context of economics research. We supplement this main data source with an array of demand and cost factors that are either specifically relevant for our setting (e.g., broadband subscriptions, data centers, internet exchange points, electricity prices) or typically used in the international trade literature to predict trade flows (as in Blum and Goldfarb (2006)). The scope of these data on both the supply and demand side allows us to estimate a detailed and flexible bandwidth demand system and a dynamic structural model of entry and competition between undersea cable operators.

Our empirical model considers the industry through the lens of a dynamic oligopoly game, with repeated Cournot competition between cable operators taking place each period in each market. We use a non-stationary Markov Perfect equilibrium concept to accommodate the evolution of exogenous demand and cost factors in our setting: our data covers much of the industry's earliest growth and witnesses substantial increases in demand and steady decreases in costs. We define a market to be a country pair and use details on cable landing points to match cables to markets. The richness of the data at hand allows us to control for many dimensions of heterogeneity, including market-level unobserved heterogeneity and regional-time shocks.

Estimation proceeds in three steps. Bandwidth demand is estimated as a function of a host of market characteristics, the number of cables serving the market, and cost-shifting instrumental variable approach to control for price endogeneity. We recover price elasticities and buyers' preference for network diversity. The demand model achieves a remarkably high level of fit, and the estimated parameters are consistent with buyers' preference for supplier diversification, with decreasing marginal returns: e.g., entry of a second cable (holding prices fixed) expands demand as much as a 23.5% decrease in bandwidth prices, whereas entry of the eighth cable is equivalent to a 6.1% decrease in prices.

Next, we recover the marginal costs of bandwidth implied by the demand estimates and the first-order conditions of the firms' period profit maximization. Third, we incorporate these profits into the dynamic game of entry and competition between cable operators and use a nested pseudo-likelihood routine (the NPL algorithm, Aguirregabiria and Mira (2007)) to estimate the dynamic investment costs and distribution of firms' private information shocks. To address concerns around convergence issues associated with the NPL algorithm (Pesendorfer and Schmidt-Dengler (2010)), we implement several alternative estimators:

e.g., two-step estimators (e.g. Pesendorfer and Schmidt-Dengler (2008)) and the spectral algorithm recently proposed by Aguirregabiria and Marcoux (2021). The latter estimator is an iterative algorithm that has the benefits of imposing equilibrium restrictions, is robust to fixed-point instability, and like the NPL algorithm, avoids the approximation of high-dimensional Jacobians. One identification challenge in this exercise stems from the nascent nature of the setting we study: we observe hundreds of instances of cable entry in our data's timeframe but only a handful of cables exiting. The lack of exit events presents us with an identification challenge which we address by leveraging the cable construction costs data to separately identify entry and fixed costs.

The estimated model is used to carry two counterfactual exercises. In the first, we examine the role of supplier diversification in shaping the dynamics of the industry, the investment in new cable systems, and the amount of surplus generated. To implement this counterfactual, we assume buyers' cannot diversify their supplier base. Entry of new cables drives competitive pressures up (and prices down) but does not benefit buyers through increased diversity. This counterfactual is equivalent to assuming perfectly correlated disruption risks across competitors in a market. We find that supplier diversification is an important determinant of entry: when buyers cannot diversify, investment in new cables is 16% lower. That is, a significant share of new capacity investment is driven by buyers' diversification needs. The net present value of total surplus created per market over the sample period is on average $895 million under the equilibrium played in the data (data generating process). Supplier diversification accounts for 13% of total surplus created and 34% of consumer surplus created.

Do market forces provide excessive or insufficient incentives for network diversity? In the second counterfactual exercise, we compare the equilibrium level of entry under the data generating process to the socially optimal level of entry. We distinguish between two sources of distortions: business-stealing effects lead to excessive entry when an entrant reduces incumbents' outputs which drives a wedge between social and private benefits of entry. Diversity effects lead to insufficient entry: the marginal entrant creates surplus by increasing supplier diversity but does not fully capture it in profits (Spence (1976), Mankiw and Whinston (1986)). Which effect dominates depend on the shape of the demand curve and the nature of post-entry competition. We separate the two sources of distortions by comparing the social planner's solution, which chooses the optimal dynamic entry path to maximizes total surplus, to a coordinated entry solution, which maximizes producer surplus (taking post-entry competition as given). In the latter scenario, only business-stealing distortions are eliminated.

We find that, relative to the market outcome, total surplus under the Planner's solution is on average 6% higher. Distortions due to diversity effects amount to between 83% and 101% of distortions due to business-stealing effects. For most markets, business-stealing tends to dominate leading to moderately excessive entry.

Our findings have implications well beyond the telecommunications industry. Supplier diversification is a key feature in the equilibrium market structure of international energy (fossil fuels or electricity) transportation markets among others. Furthermore, the findings have significant ramifications for regulators tasked with evaluating the competitiveness of an observed market structure in such an industry. A failure to account for the additional demand generated by incremental suppliers who represent an opportunity to diversify risk could lead such a regulator to incorrectly conclude that a market is sufficiently competitive. In such industries, firms' private incentives to maintain diversity would need to be accounted for when evaluating mergers.

Our findings also shed light on early capacity investment in internet infrastructure which was regarded as excessive, leading to the so-called "Telecom Crash" after the bursting of the dot-com bubble (The Economist (2002)). In the earliest days of the industry, and with the the passage of the 1996 Telecom Act, the number of fiber cables being laid (among other telecom equipment) was regarded as far too high given the amount of bandwidth demanded. A proper understanding of the role that diversification plays in mitigating disruption risk, however, adds nuance to this assessment: the market could support (indeed, would demand) more cables operating than the quantity of traffic alone would dictate, and would value the resiliency that parallel cable systems help achieve.

The rest of the paper proceeds as follows: after a literature review, Section 2 develops a simple model to illustrate how supplier diversification can shape the nature of competition. In anticipation of the empirical exercise, Sections 3 and 4 provide background on the empirical setting and details on the data used. Section 5 describes the empirical model, before Section 6 discusses our estimation strategy and results. Section 7 presents the counterfactual exercises and Section 8 concludes.

Literature Review

This paper contributes to several strands of the literature: the literature on dynamic games of entry and exit in non-standard demand settings; the literature studying endogenous product variety; the economics of the internet and its infrastructure; and somewhat less directly, the macroeconomic and trade literatures on supply chain disruptions and the propagation thereof.

The first strand of literature is the empirical industrial organization literature concerned with the estimation of dynamic entry games, especially those with complications on the structure of demand, the production technology, or the regulatory environment. For example, Collard-Wexler (2013) demonstrates that the cyclical nature of demand has significant ramifications for the firm size distribution and market structure in the ready-mix concrete industry; Kalouptsidi (2014) and Jeon (2022) study a related question in the bulk shipping industry, where the equilibrium prices that customers pay and investment decisions are significantly affected by time-to-build and demand uncertainty. This paper contributes to this literature by considering an environment with supply uncertainty and examining the impact of demand for diversification on the equilibrium market structure and welfare outcomes. Methodologically, we build on the literature on the estimation of dynamic games (Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008)). As we study a relatively nascent industry with exponential growth in demand, our treatment of non-stationarity is close to other high-tech commodity industry studies such as Igami (2017) and Igami and Uetake (2020).

This paper is also related to the empirical literature on endogenous product choice (Draganska et al. (2009), Sweeting (2013), Eizenberg (2014), Berry et al. (2016), Wollmann (2018), Fan and Yang (2020)). This literature analyzes the impact of competition on product variety and shows that similar distortions as the ones we consider can lead to too many or too few products being offered. We highlight a distinct source for preferences for product variety, i.e., supply disruptions, which has received less attention in this literature.

We take inspiration from a strand of the operations research literature, on supply chain management, that analyzes the question of how the risk of (upstream) supplier disruption or default affect competition and prices paid by downstream buyers. For example, in the model of Babich et al. (2007), a retailer places at-risk orders from suppliers before defaults events are realized; the correlation of default risk among suppliers affects the degree of differentiation and the intensity of competition. Babich et al. find that as long as there are not too many suppliers, the retailer would actually prefer to have a highly correlated default risk: the loss in supplier diversification is outweighed by the reduction in differentiation among suppliers, which leads to lower prices. Our model is closer to that of Tang and Kouvelis (2011), who study how default risk between two suppliers affects quantities and prices charged to a monopolist or duopolists engaged in downstream Cournot competition.

Third, this paper contributes to the literature studying the economics of the internet, and in particular, the infrastructure supporting it. Early theoretical contributions analyzed competition, inter-connection and peering, and antitrust matters in the (terrestrial) internet backbone market: e.g., MacKie-Mason and Varian (1994), Laffont et al. (2003), Crémer et al. (2000); Besen et al. (2001); Laffont et al. (2001); and Caillaud and Jullien (2003).

Greenstein (2015) offers a book-length treatment of the commercialization phases of the internet. In the sphere of undersea internet cables specifically, Hjort and Poulsen (2019) find that the arrival of the first undersea internet cable to various African locations has a significant positive effect on local employment, wages, and firm entry and productivity. Studying an older form of undersea communication, the early transatlantic telegraph cables, Steinwender (2018) follows in the tradition of Hoag (2006) and Garbade and Silber (1978) showing that markets on either side of the Atlantic ocean became more tightly linked with the speed of information, and the efficiency of trade flows likewise increased. Other important aspects of the internet supply chain have recently been the subject of empirical analysis: e.g., Greenstein and Pan Fang (2020) study entry strategies used by third-party data centers in the United States.

Finally, the phenomenon of disruptions examined in this paper is closely related to recent work studying global supply chain disruptions and shock propagation in the fields of international trade and macroeconomics. Recent theoretical work (Jiang et al. (2022), Baldwin and Freeman (2022), Elliott et al. (2022), Grossman et al. (2023)) studies optimal diversification policy under supply network disruptions. Supply chain disruptions can have significant costs as documented by Barrot and Sauvagnat (2016), Boehm et al. (2019), and Carvalho et al. (2021), who exploit the timing of natural disasters to estimate the role of firm-level linkages in the propagation of supply shocks. Closest to our paper, Castro-Vincenzi (2022) models how multinational automobile manufacturers endogenously reallocate production in the face of climate shocks and make investment decisions that anticipate the possibility of these shocks. We contribute to this literature by studying the endogenous response to supply disruptions through demand for diversification and how this shapes the nature of competition between suppliers and the long-run evolution of market structure.[5]

2 A Simple Example

We begin with an illustrative example demonstrating how supplier diversification can affect demand and the nature of competition, in anticipation of the detailed empirical model. A linear demand specification is used to obtain closed-form solutions and highlight the main intuition and key features that will inform the general model.

We consider the case of a buyer facing n symmetric suppliers of a homogeneous product. The baseline case without supplier disruptions is standard. The buyer maximizes her utility

[5]These topics are related to, but distinct from, the question of competition between suppliers and the decision to vertically integrate as in e.g. Loertscher and Riordan (2019); vertical integration may circumvent hold-up or market power, but would not necessarily solve the risk of disruption.

$$U(q_1,\ldots,q_n)-\sum_{i=1}^{n}p_iq_i$$

where q_i is the amount purchased from supplier i and p_i its price. The function U is assumed to be quadratic and strictly concave in the total quantity purchased

$$U(q_1,\ldots,q_n)=a\sum_{i=1}^{n}q_i-\frac{b}{2}\left(\sum_{i=1}^{n}q_i\right)^2$$

where a and b are positive constants.[6] The utility function gives rise to a linear demand structure. The inverse demand facing each supplier is given by

$$p_i(q_1,\ldots,q_n)=a-b\sum_{i=1}^{n}q_i \tag{1}$$

in the region of the quantity space where prices are positive. Because products a perfect substitutes, only total quantity matters in determining the price. An increase in any supplier's output reduces the market price—and hence the price for each supplier—by an equal amount.

Next, we introduce supply disruptions in the model above. In this case, each supplier's yield becomes uncertain because that supplier may experience a disruption. If the buyer purchases quantity q_i from supplier i, the *delivered* quantity is given by

$$\widetilde{q}_i=\delta_iq_i$$

where δ_i is a random variable representing supplier i's disruption event; it is distributed over a support $\delta_i\in[0,1]$ with mean μ and variance σ^2. The disruption risk is independently and identically distributed across suppliers. We discuss the case with correlated disruption risk below. The limit case without supplier disruptions is obtained by setting $\mu=1$ and $\sigma^2=0$.

The buyer maximizes her expected utility

$$\mathbb{E}\left[a\sum_{i=1}^{n}\widetilde{q}_i-\frac{b}{2}\left(\sum_{i=1}^{n}\widetilde{q}_i\right)^2\right]-\sum_{i=1}^{n}p_iq_i=a\mu\sum_{i=1}^{n}q_i-\frac{b}{2}\left[(\mu^2+\sigma^2)\sum_{i=1}^{n}q_i^2+\mu^2\sum_{j\neq i}q_iq_j\right]-\sum_{i=1}^{n}p_iq_i$$

With this possibility of disruption, the inverse demand facing supplier i is now given by

[6] In our application, buyers are firms located downstream in the vertical supply chain. In this case, the quadratic objective function can alternatively be thought of as a concave revenue function, e.g. $Q(a-bQ)$.

$$p_i(q_1,\ldots,q_n) = a\mu - b(\mu^2+\sigma^2)q_i - b\mu^2\sum_{j\neq i} q_j \tag{2}$$

There are two implications of supplier disruptions on demand. First, comparing Equations (1) and (2) reveals that the possibility of supplier disruptions induces horizontal differentiation across suppliers. Because $(\mu^2+\sigma^2) > \mu^2$, an increase in firm i's output has a greater (negative) effect on its price, than an increase in its rivals' output. Preference for supplier diversification makes the product ex-ante differentiated across suppliers and so insulates supplier i's demand from its rivals' choices. Regardless of the conduct model (e.g., differentiated Cournot or Bertrand), competition between suppliers is dampened due to reduced business-stealing incentives because the buyer values diversifying risk across multiple suppliers.

Second, there is a market expansionary effect from the possibility of supplier disruptions: holding prices fixed, the number of suppliers increases aggregate demand in the market. This effect can be seen by deriving aggregate demand assuming symmetric quantities and prices in equilibrium. In the baseline case without supplier disruptions, total quantity demanded is

$$Q(p) = \frac{a-p}{b}, \tag{3}$$

which is independent of the number of suppliers. With supplier disruptions, total quantity is

$$Q(p,n) = \frac{a\mu - p}{b\left(\mu^2 + \frac{\sigma^2}{n}\right)} \tag{4}$$

which is increasing in the number of suppliers n. By allowing the buyer to better diversify disruption risk, the entry of additional suppliers expands demand outward. This aspect of the model shares similarities with models where consumers value product variety (Spence (1976), Mankiw and Whinston (1986)). These implications hold for the case where suppliers' default risk is correlated.[7] We illustrate the *ceteris paribus* effects of the number of suppliers n and standard deviation σ^2 on the demand curve in this model with a numerical example

[7] Denoting the covariance across suppliers' default risk $cov(\delta_i,\delta_j) = \rho\sigma^2$, it can be shown that the inverse demand facing supplier i is

$$p_i(q_1,\ldots,q_n) = a\mu - b(\mu^2+\sigma^2)q_i - b(\mu^2+\rho\sigma^2)\sum_{j\neq i} q_j$$

The degree of horizontal differentiation is moderated by the correlation in default risk. With perfectly correlated default events ($\rho = 1$), the products are perfect substitutes across suppliers.

shown in Figure A4 in Appendix C.

In Section 5, we specify a more flexible demand model under supplier disruptions, a supply side that features static differentiated Cournot competition, and dynamic entry decisions by new suppliers. Before doing so, Sections 3 and 4 provide background information on the undersea internet cable industry and a summary of the data that motivate specific features of our empirical model.

3 Industry Background

This section provides background information on the undersea fiber-optic cable industry. Much of the information in this section is drawn from Dormon (2016) and TeleGeography (2022).

Today's fiber-optic undersea cables are the successors to a market in inter-continental communication that stretches back more than 150 years: the first transoceanic cable was a telegraph line laid across the Atlantic from 1854 to 1858 from Ireland to Canada. This line only lasted about three weeks before breaking beyond repair, but successfully demonstrated the feasibility of a transatlantic line and a stable connection was formed by 1866. The successful connection between continents allowed information to flow much more quickly and reliably, and had radical effects on a variety of industries and markets. Telegraph lines were followed over the decades by transatlantic phone lines in 1956 and transatlantic fiber-optic lines in 1988, originally used to carry phone calls (Swinhoe (2021)).

Undersea fiber-optic cables consist of several layers. In the center is a core of several glass optical fibers which do the cable's work of carrying light from one point to another. Around the core are various protective layers along with a copper layer to carry electrical power. The entire cable is typically about 25mm or 1 inch in diameter, though additional protective layers may lead to thicker portions in shallower waters. The electrical current is used to power repeaters or optical amplifiers positioned every 60-100km along the cable's length to prevent signal degradation. The power itself comes from the landing stations at either end of the cable: these stations are where the information to be carried is passed onto a cable from a datacentre or local network on one end, and in turn taken off of the cable at the other end.

The fiber-optic cables themselves are built with several fiber pairs (for inbound and outbound traffic), though cables tend to come online with only a subset of pairs "lit" or activated. Unlit pairs represent spare capacity that can be brought online without laying a new cable, though incremental investment in capital equipment at either landing station is required to light new pairs. Additional capacity can also be added on an existing lit fiber pair by using new wavelengths for data transmission.

Estimates of industry costs vary across cables, but we provide some illustrative figures here. The cost of the cable itself ranges from $10 to $25 thousand dollars per kilometer; the repeaters cost between $200 and $500 thousand; and landing stations (including branching units and other related equipment) can cost from $1 to more than $10 million, depending on the size and complexity. Total constructions costs are on average in the hundreds of millions of dollars. Ongoing operational costs include electricity costs and landing centre staffing and maintenance costs in the range of $5 million or more annually, and cable repairs.

Undersea cable failures occur with surprising frequency, most commonly from accidental shipping interference (e.g. fishing trawlers or anchors) though interruptions from natural causes and intentional sabotage also occur.[8] The rate of failure is significantly lower than terrestrial cables, though it is offset by a much higher cost and longer length of repair: estimates range from $250 thousand to $1 million, with a time frame of one to several weeks. Industry estimates are that 100 to 200 submarine cable failures occur each year; demand for repairs to the roughly 400 active cables (in addition to the laying of new cables) is high enough to sustain a global fleet of at least 60 cableships (ISCPC (2022)).

Today, the industry continues to grow in size and recognized importance. The quantity of used international bandwidth has seen an astonishing 49% annual growth rate between 2005 to 2021. Significant investment forecasted over the coming years indicates that the industry's growth is expected to continue: the more than 400 cables in service are set to be joined by at least another 43 cables by the end of 2026 at an estimated cost of $39.5 billion.[9]

Cables are designed with a theoretical working lifespan of 20 to 30 years. Because the industry is recent, only a few instances of major cable exits are observed in our data and occur in majority near the end of our sample—after 2020.[10] Whether the first entrants will start a small wave of cable retirements in the coming years, or cables will operate past their designed lifespan in the face of rising demand remains to be seen.

Current ownership of undersea fiber-optic cables is unconcentrated at the global level: 400 unique owners (which are referred to as "network service providers") are reported for the 434 cables in our data. The median cable is owned by a single owner, and the mean number of owners per cable is 2.5. The median owner in our data has an ownership share in only a single cable, and the mean owner has a share in 2.9 cables; when considered on a per-market basis, the median remains at 1 while the mean drops to 1.1. However, the number of cables in an

[8] Defense experts have long been aware of the threat posed to this infrastructure, and public awareness is increasing after incidents like the Nord Stream gas pipeline sabotage in September 2022.

[9] Source is authors' calculations from TeleGeography datasets. More details are found in Section 4 and Tables 1 and 2.

[10] Decommissioned cables include TAT-14, CANTAT-3, Tasman 2, Atlantis-2.

owner's portfolio has a skewed distribution: 23 owners (major telecommunications companies and internet firms) have an ownership stake in at least 10 cables. Outside of these players, owners tend to be governments and smaller regional telecommunications firms. Tables A4 and A5 in Appendix C provide more details on cable ownership.

Bandwidth on these cables is sold in wholesale markets and is purchased by a variety of customers, which can be broken down into four broad categories: internet service provider (ISP), content provider, private enterprise, and research institutions/other.[11] ISPs are the traditional telecommunications companies that businesses and consumers rely on to connect to the internet; some of these firms own cables, and others lease capacity on others' cables (some do both). Content providers include media companies as well as technology firms with large bandwidth needs such as Meta, Google, Amazon; unsurprisingly this category has represented a growing share of global bandwidth demand in recent years. The final two categories of private enterprise and research/educational customers represent the two smallest sources of demand.

The only competitor to undersea cables for transmitting data intercontinentally is the recent advent of networked data satellites. Some of these networks are being created by high-profile firms, such as SpaceX's Starlink, or Amazon's Kuiper. However, these networks are properly viewed as complements rather than competitors to undersea cables. For example, the entire Starlink network is reported to have a capacity of 10.3 Tbps and to have cost roughly $30 billion to deploy. In contrast, the recent Dunant cable built between Europe and North American is reported to have cost $165 million and has a capacity of 250 Tbps. Given this cost disparity relative to undersea cables, satellites are more properly understood as a means to expand high bandwidth internet access to remote or sparsely populated regions; traffic from those satellite networks will typically be sent to ground stations and end up travelling over terrestrial or undersea networks.

The next section details the data sources used in this study, and provides relevant summary statistics. Section 5 follows and develops a formal model which accounts for the salient industry features discussed here.

[11] Bandwidth is a product that is vertically differentiated by speed, i.e., buyers can purchase leases at 10Gbps, 40Gbps, 100Gbps, etc. Over the period studied in this paper, 10Gbps is by far the most common product purchased (100Gbps was introduced only in the last few years of our sample and 40Gbps is relatively rare), as discussed in the next section. We focus on the market for 10Gbps and treat the product as homogenous.

4 Data and Descriptive Statistics

This section describes the data used in our empirical analysis: its sources, the definitions of variables of interest, and key trends in those variables. The primary data, on undersea cables and bandwidth are from industry data provider TeleGeography. We complement this source with supplementary data on demand and cost factors from various auxiliary sources.

4.1 TeleGeography

The bulk of the data used in this study is from data provider TeleGeography, a telecommunications market research firm which has been in operation since 1989. TeleGeography collects and retains comprehensive datasets on various aspects of the telecommunications industry. We make use of data from two separate datasets focusing on undersea fiber-optic telecommunications: The Global Bandwidth Research Service and the Wavelengths Pricing Suite.

The Global Bandwidth Research Service (GBRS) contains detailed data on over 400 undersea cable systems. The data on each cable system includes: landing points, date of entry into service, ownership, designed (or potential) capacity, length, and construction costs.[12] In addition, GBRS features a yearly panel of used bandwidth at the country-pair level and data on the roll-out of cloud data centers owned by major technology firms.[13] Used bandwidth describes the sum of bandwidth deployed actively in a route and is not idle, it provides a measure of underlying international bandwidth demand. We present summary statistics from the GBRS data in Tables 1 and 2.

Table 1 contains descriptive statistics on some of the data series in the GBRS data pertaining to undersea cable supply. The table contains data for the global network of undersea cables. Columns 2 and 3 present the number and total length (in km) of new cables that come online in each year; cumulative counts of the number of active cables and the total network length are shown in columns 5 and 6. Column 4 contains the (unweighted) average cost of the new cables in each year, normalized by the distance of the cable; costs are reported for only 47% of cables in the data. Columns 7 through 9 contain information on potential and active capacities for network: potential capacity (in Tbps) is shown in column 7, and lit (or active) capacity is reported in column 8. The share of capacity that is active (i.e., the utilization rate) is reported in column 9.

[12] A limitation of these data is the absence of information on lit capacity over time at the cable level, which prevents us from calculating cable-level market shares.

[13] These are Meta, Microsoft, Amazon, and Google. The count data we use corresponds to the number of "cloud regions." A cloud region is a physical location (e.g., Palo Alto, US) where providers typically cluster multiple data centers (e.g., AWS operates three data centers in the Palo Alto cloud region).

Several trends are observable from Table 1. The industry has experienced a robust rate of growth: the number of active cables and the network's total mileage have more than doubled over this time frame. This growth is even more dramatic when one takes into account the larger capacity of more recent cables: both potential and lit capacity have increased by two orders of magnitude. One final observation is that the entry rate has been steady despite a significant amount of spare, or unlit capacity in each year – the share of capacity that is lit peaks at only 35% in 2020. This suggests that capacity constraints are not a first-order concern in this industry over our sample period.

Table 1: Descriptive Statistics on Undersea Cable Supply

Year	New Cables	New km '000s	Mean cost \$m / 1000km	Cum. Cables	Cum. km '000s	Potential Tbps	Lit Tbps	Share Lit
2005	13	24.9	44.3	180	572.5	87.1	10.6	0.12
2006	11	31.7	36.5	191	604.2	89.2	14.3	0.16
2007	11	7.2	80.7	202	611.4	100.5	16.5	0.16
2008	20	44.3	54.6	222	655.7	121.0	26.8	0.22
2009	17	69.1	44.0	239	724.8	142.3	34.3	0.24
2010	14	61.0	41.5	253	785.7	239.5	44.1	0.18
2011	14	29.4	39.3	267	815.1	309.8	54.5	0.18
2012	22	61.6	217.4	289	876.7	470.7	73.5	0.16
2013	11	16.9	56.6	300	893.6	748.4	93.4	0.12
2014	15	30.2	141.9	315	923.9	981.2	137.4	0.14
2015	9	19.3	61.0	324	943.1	1174.6	193.9	0.17
2016	16	64.7	466.6	340	1007.8	1479.7	292.4	0.20
2017	15	74.1	107.8	355	1081.9	1836.6	412.0	0.22
2018	18	74.0	34.7	373	1155.9	2359.3	563.8	0.24
2019	23	31.5	261.2	396	1187.4	2521.5	784.6	0.31
2020	21	68.6	103.7	417	1256.0	3171.1	1095.2	0.35
2021	17	45.3	225.3	434	1301.3	3928.3	1346.7	0.34

We report further summary statistics in Table 2, turning attention now to factors of demand. Columns 2 and 3 report the number of unique cities and countries connected to the internet via an undersea cable by year. Column 4 reports the total used (international) bandwidth by year, in Tbps. Columns 5 and 6 provide information on the data centres present in TeleGeography's dataset: column 5 reports the number of new data centers built in each year (beginning in 2006), and column 6 reports the total cumulative count. Dramatic growth in the industry is also evident from Table 2. The number of cities connected directly to an undersea cable more than doubles, while the number of countries increases from roughly 66% to nearly 90% of all countries. The number of data centers increases from 0 to 118, with much of that growth coming in the last several years. Finally, the growth in used bandwidth is a stark three orders of magnitude over this time frame.

We also source data on bandwidth prices from TeleGeography's Wavelengths Pricing

Suite. The Wavelengths dataset contains quarterly prices collected by TeleGeography from bandwidth providers. Each observation in the dataset represents the quoted monthly price on a 1-year unprotected lease in a particular capacity segment (e.g., 10Gbps, 100Gbps) on a particular long-haul city-to-city route from a particular bandwidth provider.[14] The provider name is anonymized. Coverage is not comprehensive, but TeleGeography reports that data is sourced on a voluntary basis from dozens of providers. In this paper, we focus on the 10Gbps segment as it is by far the most commonly bought capacity over our sample period.

Table 3 presents descriptive statistics from this dataset for the 10Gbps market segment from 2005 to 2021. Column 2 presents the number of unique price quotes each year, and columns 3-5 contain the 25th, 50th and 75th percentile of quoted prices (in thousands of US$). Table 3 reveals a few empirical features worth noting. First, prices fall significantly over time at all reported percentiles. Second, significant price dispersion exists within a segment-year, with the ratio of 75th to 25th percentile prices ranging from roughly 2 to more than 10. This price dispersion is expected given the heterogeneity in costs and distances across markets (e.g., U.S.-Japan versus U.K.-France). Third, the rise in the number of cities connected by cables (shown in Table 2) is reflected in the increase in the number of price quotes available.

Figure 2 provides some initial descriptive evidence for the relationship between bandwidth used and prices—abstracting from issues of endogeneity of prices and market structure. Panel (a) plots price (measured as log of price / distance) against quantity (measured as log of used bandwidth) by market-quarter, broken up by the number of cables operative in that market-quarter. Markets are defined as country pairs and we provide more details about the market definition in Section 4.3. Each sub-panel also plots the line of best fit. Panel (b) combines these lines of best fit into a single plot, wherein it can be seen that the market-quarters with a higher number of cables on average feature a higher price/distance conditional on used bandwidth.

4.2 Supplementary Data

In addition to the data presented above, we make use of several auxiliary data sources in our estimation approach.

A first set of auxiliary data is sourced from the CEPII Gravity Database, a repository of macroeconomic and trade-related variables that are suitable for the estimation of the determinants of international trade (Conte et al., 2021). We source data on annual GDP figures,

[14]Protected bandwidth refers to a leased capacity on an undersea internet cable that is backed up by a secondary, "protection" cable. In the event of a failure or outage on the primary cable, traffic can be automatically rerouted to the secondary cable, minimizing downtime. Unprotected bandwidth, on the other hand, does not have this secondary backup and is therefore at a higher risk of outage. Unprotected bandwidth is by far the most traded type of bandwidth on undersea cables.

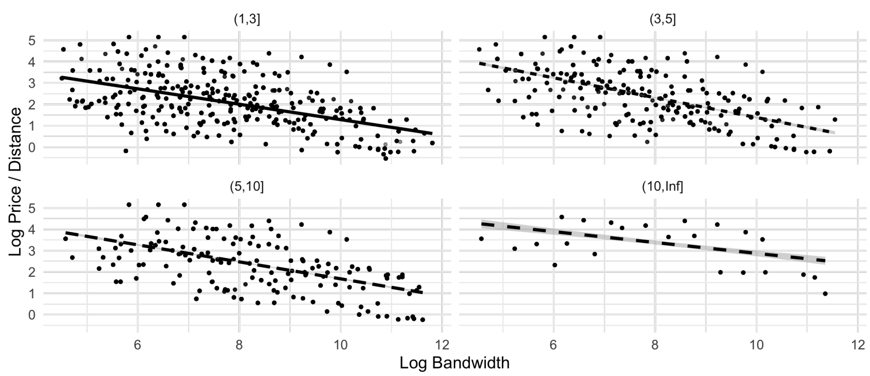
Table 2: Descriptive Statistics on Bandwidth Demand

Year	Cities	Countries	Used Bandwidth	Datacenters New	Datacenters Cumulative
2005	549	126	5.0		
2006	603	131	7.0	2	2
2007	623	132	11.5	3	5
2008	665	133	19.2	1	6
2009	694	141	30.6	4	10
2010	726	148	46.8	4	14
2011	753	150	70.1	7	21
2012	820	164	101.2	6	27
2013	834	165	145.7	2	29
2014	890	167	212.5	10	39
2015	923	167	300.0	6	45
2016	949	167	443.0	12	57
2017	984	169	666.6	13	70
2018	1,040	172	997.4	12	82
2019	1,097	172	1,466.4	13	95
2020	1,152	175	2,124.8	15	110
2021	1,189	175	2,885.6	8	118

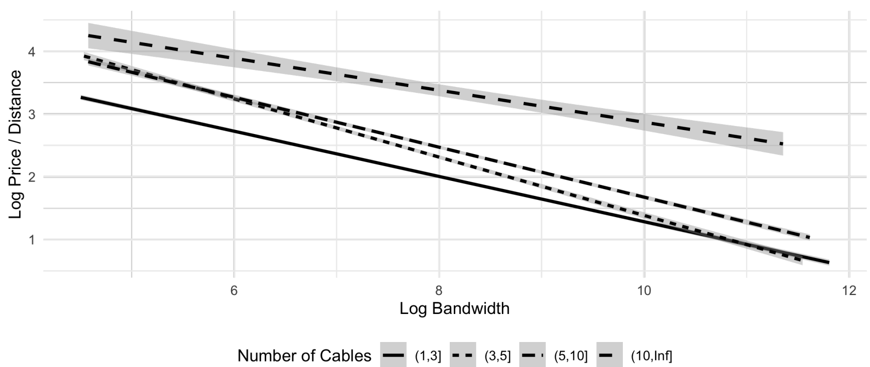
Table 3: Descriptive Statistics on Bandwidth Prices (in thousands of US$)

Year	Postings	10 Gbps Price Percentiles 25th percentile	50th percentile	75th percentile
2005	359	31.8	55.5	75.9
2006	867	13.5	22.0	38.3
2007	1,437	12.0	15.0	23.3
2008	2,510	10.0	13.0	18.0
2009	1,681	8.9	11.9	17.1
2010	1,830	7.2	10.9	19.8
2011	2,334	5.5	9.4	45.0
2012	2,431	5.0	9.0	50.0
2013	2,531	4.1	6.6	42.1
2014	2,555	3.3	6.0	45.0
2015	2,512	2.6	5.0	30.0
2016	2,967	2.1	4.5	28.0
2017	3,547	1.1	3.2	24.0
2018	3,623	1.1	3.1	19.9
2019	3,602	1.2	2.1	16.2
2020	3,124	1.3	2.4	15.0
2021	3,309	1.1	2.2	12.0

Figure 2: Empirical Relationships Between Price and Quantity



(a) Log Price / Distance and Log Quantity by Number of Cables



(b) Combined Lines of Best Fit

Notes: Panel (a) plots the log of price / distance against the log of used bandwidth for each market-quarter, split up by the number of cables active in each market, along with a line of best fit. Panel (b) combines the lines of best fit. A few outlier observations with low levels of used bandwidth are not shown on this graph for the purpose of enhancing clarity and visualization.

bilateral trade flows between countries, distance between countries, as well as information about whether two countries share a common official language or land border.

Another data series used in estimating demand is the number of broadband subscriptions, which we take as a proxy for end-point internet penetration. We access this data from the World Bank, which maintains an annual panel at the country level tracking the number of fixed broadband subscriptions with speeds greater than or equal to 256Kb per second. This includes both residential and business customers, but one limitation is that the focus on fixed subscriptions means that the number of mobile phones capable of accessing the internet are missed.

When estimating demand for bandwidth, we make use of an instrumental variable strategy where the instruments are related to electricity prices (input costs) and so can be thought of as classical cost-shifters. We use panels of electricity generation shares from Our World in Data (ourworldindata.org). Detailed data exists on country-level shares of electricity generated by coal, gas, and fossil fuels (inclusive of coal and gas); we infer the oil share by subtracting coal and gas shares from fossil fuel shares. We then download price series for coal, gas, and oil from the Federal Reserve Economic Data (FRED) (https://fred.stlouisfed.org/). We use the Brent crude series for oil, the global Australian coal price for coal, and the global EU price of natural gas; the latter two prices are converted to dollars per barrel of oil equivalent (BOE).

Finally, we compile data on cable disruptions and faults. TeleGeography's GBRS provides data on cable faults: they report publicly disclosed faults as early as 2016 with detailed information (e.g., date fault discovered, start of repairs, length of repairs, cause). We supplement this by hand-collecting data on faults for 2013-2016 from a selection of newspapers, and end up with a final sample of 168 faults. Of these 168 faults, data on the duration of repairs is available for 115 (or 68%); the median time to repair a fault in our data is 13 days, while the 10th and 90th percentile of repair times are 1 and 55 days respectively. Details on the number of faults in this dataset by year is presented in Table A6.

The number of faults for which we are able to gather data is an under-count of the actual fault total we should expect according to industry experts; it is clear that this data collection methodology selects for only "high-profile" faults of major cables or publicly disclosed ones. This data selection issue prevents us from using this cable fault data as an integral part of our identification strategy. However, we do leverage the cable faults data in our estimation approach to verify that our estimated cable operating costs are consistent with the cable fault propensity. We also exploit the faults data, in robustness checks, as an exogenous shifter of entry costs (and subsequent market structure) in the demand estimation.

4.3 Market Definition

In anticipation of the industry model presented in the next section, we discuss our market definition. As outlined above, our primary data for estimating demand comes from TeleGeography's GBRS and and Wavelengths Pricing Suite.

The GBRS dataset contains data on used bandwidth at a country-to-country level on an annual basis. The Wavelength Pricing dataset contains price quotes on bandwidth at a city-to-city level on a quarterly basis. To strike a balance between the granularity of the pricing data relative to the used bandwidth data, we combine the two datasets as follows. First, we aggregate price quotes from the city-to-city level to the country-to-country level (we use a weighted-average specification where the weights are the number of city-to-city price quotes, but have also experimented using simple average and median); we also multiply the quoted monthly lease prices by three to convert them from a monthly to a quarterly price. Second, we interpolate the annual bandwidth used to the quarterly level.

At the end of this process, a market-period in our estimation sample is a country-to-country in a calendar quarter. Unsurprisingly this panel is not perfectly balanced: new cables enter and price quotes are collected for new country-to-country pairs over time. Details on the sample of markets used are presented in Appendix A.[15]

5 Industry Model

This section presents the dynamic structural model. The model will be subsequently used to guide the estimation and recovery of cable operators' period profits and dynamic investment costs. In turn, we will use these estimates in two counterfactual simulations. The first counterfactual assesses the role that the demand for network diversity plays in driving entry and surplus creation; the second evaluates the efficiency of entry outcomes and quantifies the various sources of distortions relative to the socially optimal level.

Cable operators' entry choices are modeled as a dynamic game played in each market. Undersea cables are durable equipment, making the decision to build a cable a typical case of investment under uncertainty (Dixit et al. (1994)). Therefore our model must account for the fact that firms may delay their entry in anticipation of more favorable demand or cost states. The central part of the model specifies how firms make their entry decisions as a function of market structure (i.e., the number of incumbent cables), and market-level

[15] Some markets contain contiguous countries (e.g., France-Spain, US-Canada). To alleviate the concern that terrestrial cables may carry a significant proportion of bandwidth in these markets, we perform our empirical analysis by either controlling for whether a market contains contiguous countries (when estimating demand) or by excluding the 34 markets with contiguous countries.

demand and cost state variables. Finally, an equilibrium of the dynamic game is specified. In this setting, the processes governing the demand and cost states are nonstationary, hence, we characterize a nonstationary Markov Perfect equilibrium of the game, where strategies and transition functions are indexed by time.

5.1 Timing

Time is discrete with an infinite horizon $t = 0, 1, 2, ...$, with a period corresponding to a calendar quarter. Markets are assumed to be independent of each other. In what follows, we consider a specific market m.

Players and States. A firm is an undersea cable operator. A finite number N of symmetric firms are indexed by $i \in \{1, \dots, N\}$. In any period t, each firm is either active in the market or a potential entrant. A firm is defined as active if it owns and operates an undersea cable.[16] We denote its state by $s_{it} \in \{0, 1\}$, where s_{it} equals 1 if the firm is active and zero otherwise. The industry state is the aggregation of firm states $s_t \equiv \{s_{it}\}_{i \in N}$ and we let $s_{-it} = \{s_{jt}\}_{j \neq i}$ denote the state of firms other than i, and $n_{mt} = \sum_{i=1}^{N} s_{it}$ the total number of firms operating in the market.

The exogenous characteristics of the market in period t consist of two main components: (1) an aggregate demand state, denoted d_{mt}, which captures various demand factors such as the level internet penetration or the number of data centers operating in the market; (2) an aggregate cost state h_{mt}, which captures supply factors such as input costs (e.g., electricity), improvements in bandwidth capacity provisioning (through technological advances), the physical and geographic characteristics of the market that determine the required cable length and number of repeaters, or the frequency of cable faults. Demand factors change over time due to population and economic growth and increasing internet access and digitization. Electricity costs evolve over time due to changes in global energy prices (e.g., of oil, gas, and coal) and country-specific changes in their electricity-generation fuel mix.

The vector of public information variables includes the industry state and exogenous market characteristics. All these variables are publicly observed and denoted by the vector $\mathcal{M}_t = (n_{mt}, d_{mt}, h_{mt})$.

Actions. Every period t, firms decide simultaneously but independently whether to be active or not in the market. Let $a_{it} \in \{0, 1\}$ be the binary indicator of the firm i's decision at period t, such that $a_{it} = 1$ if the an incumbent firm decides to remain active in the market at the end of period t or a potential entrant decides to enter, and $a_{it} = 0$ if an incumbent exits

[16] We rule out ownership of multiple cables by the same firm in a given market as this is rare in practice (see discussion of cable ownership in Section 3).

or a potential entrant stays out at the end of period t. Firms that exit or potential entrants that decide to stay out are replaced by a new set of potential entrants in the following period. We use the variable a_{-it} to denote the vector of actions taken in period t by all firms except firm i.

Firms' choices are dynamic because of partial irreversibility in the decision to enter a market, i.e., sunk costs. At the end of period t, firms simultaneously choose their action a_{it} which determines their next-period state s_{it+1} with an understanding that their choice will affect their variable profits in subsequent periods. We model the choice of entry and exit as a game of incomplete information, so that each firm i has to form beliefs about other firms' entry and exit choices. More specifically, there are components of the entry costs and profits of a firm which are firm-specific and constitute private information.

5.2 Period Profits

Firm i's period profits, net of private information shocks, are

$$\pi_{it}(a_{it}, \mathcal{M}_t) = VP_i(\mathcal{M}_t) - FC_{it}(\mathcal{M}_t) - EC_{mt}(s_{it}, a_{it}) + EV_{mt}(s_{it}, a_{it}), \tag{5}$$

where $VP_i(\mathcal{M}_t)$ are variable profits, FC_{it} is the fixed cost incurred by firm i to operate an undersea cable, EC_{it} is the entry or set-up cost of a new cable, and EV_{it} is the scrap value from retiring an exiting cable.

We distinguish two main components in firms' profit at time t: variable profits and fixed profits (or costs). The variable profits $VP_i(\mathcal{M}_t)$ are equal to the difference between revenue from selling bandwidth on the undersea cable and the variable costs of operating the cable (e.g., maintenance, energy costs). It varies continuously with the firm's output and it is equal to zero when output is zero. Variable profits depend on the demand and supply states (d_{mt}, h_{mt}) as well as the number of rivals competing in period t n_{mt}. Variable profits are received by any firm who is active at the beginning of period t ($s_{it} = 1$).

Bandwidth is a high-tech commodity with limited scope for differentiation, and hence we assume Cournot competition among undersea cable carriers that are active in period t. We focus on a Nash equilibrium in the spot market for bandwidth. Thus, the market structure (summarized by the industry state s_t), along with (d_{mt}, h_{mt}), completely determines each firm's equilibrium variable profit $VP_i(\mathcal{M}_t)$ from competing in period t. This parsimonious formulation allows us to handle the dynamic oligopoly game of entry and exit with a tractable state space.

The fixed part of period payoff derives from buying, selling, or renting inputs that are fixed during the active life of the cable and that are necessary for the firm to operate in the market. These fixed inputs may include cable equipment, repeaters, landing stations, or even managerial skills. We distinguish three components in the fixed payoff. The fixed cost of an active firm, $FC_{it}(\mathcal{M}_t)$, is incurred in any period where the firm is active ($s_{it} = 1$). Fixed costs reflects the need for continual investment in maintenance and upgrades to facilities and equipment. The entry cost of a new entrant, $EC_{mt}(s_{it}, a_{it})$, is paid if the firm is a potential entrant ($s_{it} = 0$) and decides to enter ($a_{it} = 1$). The scrap value of an exiting cable, $EV_{im}(s_{it}, a_{it})$ is received by the firm if the firm is an incumbent ($s_{it} = 1$) that exits in period t ($a_{it} = 0$).

At the beginning of period t, each firm draws a vector of private information shocks associated with each possible action $\boldsymbol{\epsilon}_{it} = \{\epsilon_{it}(a)\}_{a\in\{0,1\}}$. We assume that the shocks ϵ_{it} are independently distributed across firms and over time and have a cumulative distribution function $G(.)$ that is strictly increasing and continuously differentiable with respect to the Lebesgue measure. These two assumptions allow for a broad range of specifications for the ϵ_{it}. In our application, these shocks will be distributed Type 1 extreme value, scaled by a parameter θ_m^ϵ which can vary by market.[17]

It will be convenient to distinguish two additive components in the period profit function:

$$\Pi_{it}(a_{it}, \mathcal{M}_t, \boldsymbol{\epsilon}_{it}) = \pi_i(a_{it}, \mathcal{M}_t) + \epsilon_{it}(a_{it}). \tag{6}$$

5.3 Dynamic Optimization and Equilibrium

Firms make their dynamic discrete choices of entry and exit to maximize their discounted sum of expected profits. They discount their future stream of profits by a factor $\beta \in (0, 1)$, with rational expectations regarding the endogenous evolution of market structure and the exogenous evolution of demand and production costs.

We focus on Markov-Perfect Bayesian Nash Equilibria (MPE). We first define firm strategies, value functions, and then the equilibrium conditions. A firm's strategy, at time t, depends only on its payoff relevant state variables $(\mathcal{M}_t, \boldsymbol{\epsilon}_{it})$. A strategy profile is denoted

$$\alpha = \{\alpha_{it}(\mathcal{M}_t, \boldsymbol{\epsilon}_{it}))\}_{i\in I, t\geq 0}.$$

Given strategy profile α, firm i's value function satisfies

$$V_{i,t}^\alpha(\mathcal{M}_t, \boldsymbol{\epsilon}_{it}) = \max_{a_{it}\in\{0,1\}} \left\{ v_{i,t}^\alpha(a_{it}, \mathcal{M}_t) + \epsilon_{it}(a_{it}) \right\} \tag{7}$$

[17]These shocks can be thought of as representing each firm's idiosyncratic conditions in terms of information, financial condition, and other transient organizational conditions that are relevant for cable investment.

where $v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_t)$ are choice-specific value functions.

The choice-specific value function for active firms are given by

$$v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_t) = \begin{cases} \pi_i(a_{it}, \mathcal{M}_t) + \beta\, \mathbb{E}_t \left[V^{\alpha}_{i,t+1}\left(\mathcal{M}_{t+1}, \boldsymbol{\epsilon}_{i,t+1}\right)\right] & \text{if} \quad a_{it} = 1 \\ \pi_i(a_{it}, \mathcal{M}_t) + EV_{mt} & \text{if} \quad a_{it} = 0 \end{cases} \tag{8}$$

where the next-period state $\mathcal{M}_{t+1}$ is formed of the next-period market structure s_{t+1}, and exogenous market-level variables $(d_{m,t+1}, h_{m,t+1})$. The distribution over next-period states is given by the transition probabilities $f_t(d_{m,t+1}, h_{m,t+1}|d_{mt}, h_{mt})$ of exogenous states which is indexed by time because processes are nonstationary, and the distribution of other firms' shocks $\Pi_{j\neq i} g(\boldsymbol{\epsilon}_{j,t})$ and strategies α_j for $j \neq i$. For potential entrants, the choice-specific value functions are given by

$$v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_t) = \begin{cases} -EC_{mt} + \beta\, \mathbb{E}_t \left[V^{\alpha}_{i,t+1}\left(\mathcal{M}_{t+1}, \boldsymbol{\epsilon}_{i,t+1}\right)\right] & \text{if} \quad a_{it} = 1 \\ 0 & \text{if} \quad a_{it} = 0 \end{cases} \tag{9}$$

A MPE is a strategy profile α^* such that for every player, state, and period

$$\alpha^*_{i,t}(\mathcal{M}_t, \epsilon_{it}) = \arg \max_{a_{it}\in\{0,1\}} \left\{ v^{\alpha^*}_{i,t}(a_{it}, \mathcal{M}_t) + \epsilon_{it}(a_{it}) \right\} \tag{10}$$

The probability that firm i chooses action a_{it} in period t given state $\mathcal{M}_t$ (hereafter, the conditional choice probability or CCP) is defined as

$$P^{\alpha}_t(a_{it}|\mathcal{M}_t) = \Pr(\alpha_{i,t}(\mathcal{M}_t, \epsilon_{it}) = a_{it}|\mathcal{M}_t) \tag{11}$$

One can express the choice-specific value function as a function of CCPs instead of strategies. That is,

$$v^{\mathbf{P}}_{i,t}(a_{it}, \mathcal{M}_t) = \pi_i(a_{it}, \mathcal{M}_t) + \beta \sum_{a_{-it}} \int \overline{V}^{\mathbf{P}}_{i,t+1}\left(\mathcal{M}_{t+1}\right) dF_t(\mathcal{M}_{t+1}|\mathcal{M}_t, a_t) P_t(a_{-it}|\mathcal{M}_t) \tag{12}$$

where $a_t = (a_{it}, a_{-it})$ and $\overline{V}^{\mathbf{P}}_{i,t}$ is the *ex-ante* value function expressed before the realization of the private shock $\boldsymbol{\epsilon}_{it}$

$$\overline{V}^{\mathbf{P}}_{i,t}(\mathcal{M}_t) = \int \max_{a_{it}\in\{0,1\}} \Bigg\{ \pi_i(a_{it}, \mathcal{M}_t) + \epsilon_{it}(a_{it}) \\ + \beta \sum_{a_{-it}} \int \overline{V}^{\mathbf{P}}_{i,t+1}(\mathcal{M}_{t+1})\, dF_t(\mathcal{M}_{t+1}|\mathcal{M}_t, a_t) P_t(a_{-it}|\mathcal{M}_t) \Bigg\} dG(\epsilon_{it}). \tag{13}$$

If private shocks are distributed Type 1 extreme value (with scale parameter θ^{ϵ}_{m}), an optimal strategy for firm i will map into equilibrium conditional choice probabilities of the form

$$P_t(a_{it}|\mathcal{M}_t, \mathbf{P}) = \frac{\exp\left(\frac{v^{\mathbf{P}}_{i,t}(a_{it},\mathcal{M}_t)}{\theta^{\epsilon}_{m}}\right)}{\sum_{a'\in\{0,1\}} \exp\left(\frac{v^{\mathbf{P}}_{i,t}(a',\mathcal{M}_t)}{\theta^{\epsilon}_{m}}\right)}. \tag{14}$$

There can be multiple equilibria. In our estimation approach, we assume that the same equilibrium is played in all markets and we verify that estimates obtained using a two-step estimator–a procedure that is robust to multiplicity—are close to those obtained using iterative methods. For our first counterfactual exercise, we initialize the algorithm at a large number of starting values and converge systematically to a unique fixed point. For the second counterfactual exercise, we solve a social planner problem (as a single-agent dynamic decision problem).

6 Identification and Estimation Approach

6.1 Identification

We follow the literature on the identification of dynamic decision problems (Rust (1994), Magnac and Thesmar (2002), Bajari et al. (2015)) and assume that the discount factor and the distribution of firm shocks (β, G) are known.[18] We allow the scale of the firm shocks to depend on the size of entry costs as described in more details below. In this setting, the scale of the firm-specific shocks is identified because variable profits are treated as observed when estimating the dynamic model.

Aguirregabiria and Suzuki (2014) study the identification of market entry and exit games. They show that the level of fixed costs, entry costs and exit value are not separately iden-

[18]In the estimation approach, we experiment with discount factors ranging from 0.90 to 0.975, with 0.95 as a baseline, and assume that firm-specific shocks follow a type-1 extreme value distribution.

tified.[19] When estimating the model, we normalize the exit value to zero and estimate entry costs and fixed operation costs. Industry reports suggest that, while raw materials and equipment from retired cables can be reused, the vast majority (94%) of unused undersea *telephone* cables are abandoned on the seabed (The Guardian (2016)), providing some support for our normalization of scrap values.

The relatively recent nature of the industry presents us with an additional challenge in that exit events are rare. Without observing exit events, we cannot separately identify entry costs from fixed costs, even under the "normalization" introduced above. To address this problem, we leverage cable-level construction costs data to estimate entry costs outside of the dynamic model. Once entry costs are estimated, we use firms' optimal entry decisions to recover their fixed costs.

The two key parameters entering our estimation and counterfactual analysis are the elasticity of bandwidth demand with respect to prices and the dependence of bandwidth demand on the number of operating cables (or network diversity). The first is identified using an instrumental variable approach based on cost shifters (energy prices). Section 6.3 provides details. The second parameter is identified under similar assumptions as in the endogenous product variety literature (e.g., Eizenberg (2014), Wollmann (2018), Fan and Yang (2020)): that is, we assume that firms make their entry decisions before the current-period transient demand and cost shock are realized. This is a natural assumption in our setting given that entry decisions are often made at least a year or more before the cable is ready for service and that timing of entry, after conditioning on state variables, comes from idiosyncratic shocks (e.g., delays in construction) that are unrelated to unobserved transient demand shocks. We provide robustness checks to this assumption in Section 6.3. We discuss the role of time-to-build in Appendix B.1.

Finally, given the availability of a long panel, we control for persistent market-level unobserved heterogeneity in demand and costs with fixed effects.

6.2 Solution Method and Estimation Approach

In our empirical approach, we follow three steps. First, we start by estimating demand for bandwidth in each market to recover price elasticities and how demand depends on the number of cables under operation (market expansion effect). Second, we recover the marginal costs of bandwidth implied by the demand estimates and the first-order conditions of the firms' period profit maximization. These static estimates for each market-period allow us to compute the variable profits per cable under different market structures and time periods.

[19] More recent contributions include Kalouptsidi et al. (2021b) and Kalouptsidi et al. (2021a).

We also estimate the transition processes of the exogenous components entering the demand and cost states. Third, we incorporate these profits into a dynamic game of entry and competition between cable operators, which we solve to estimate the fixed costs of operating a cable and the scale of the firm-specific logit shocks.

In this section, we take the outputs of the first two steps as given and outline the methods used to solve and estimate the dynamic game (third step). We begin by discussing how the model is solved, as estimation of the dynamic model involves only small extensions to the solution procedure.

Solution Method. The dynamic game is solved via policy iteration (Judd (1998), Rust (2000)). This approach consists in iterating repeatedly between two steps: a given iteration starts by updating the ex-ante and choice-specific value functions given the current vector of CCPs (policy evaluation), then these value functions are used to update the vector of CCPs (policy improvement). The algorithm iterates until value functions and CCPs converge, up to a pre-defined tolerance level.

Three important features are worth highlighting. First, because the demand and costs states may follow non-stationary processes, we adopt as equilibrium concept a symmetric non-stationary MPE, in which time becomes a state variable. To maintain tractability, we assume that the industry enters a stationary regime after some period T (in practice, we use the last quarter of 2021). Second, because there are too few exit events over our sample period, we do not model exit decisions. Cable operators choose optimally when to enter, but once in the market, we assume that incumbents remain active. We return to this assumption and incorporate cable design life (in the order of 25 years) and exogenous retirement dates in Appendix B.2. Third, we assume that, among all potential entrants, only one firm has an opportunity to enter each period.[20] This assumption serves two purposes: it rules out the possibility of multiple cables entering in the same period which does not occur in our data; and as detailed in Section 7, it greatly simplifies the solution of the social planner's problem.

The dynamic game is solved by backward induction starting from the last period in our sample, i.e., $t = T$, market by market. In market m, denote a given element of the state space as

$$\mathcal{M}_{m,t} = (n_{mt}, d_{mt}, h_{mt})$$

where n_{mt} is the number of cables in operation, d_{mt} is the exogenous demand state, and h_{mt} is the exogenous cost state. The latter two variables are discretized for tractability and we provide details about the discretization procedure in the next two sections. To account

[20] The total number of firms N per market is set to the maximum number of cables observed in the data for that market plus two.

for market-level unobservables and the non-stationary nature of the industry, the vector of CCPs and value functions are indexed by the market and the period (in addition to the state).

We iterate over the following steps.

1. Initialize the vectors of CCPs for each market m, state $\mathcal{M}_{m,t}$, and period t, denoted $\mathbf{P}_{m,t}$. The vector $\mathbf{P}_{m,t}$ is indexed by the state $\mathcal{M}_{m,t}$ and gives the CCP of entry into the market m in state $\mathcal{M}_{m,t}$ in period t.

2. For each market m and period t, form the transition matrix from state $\mathcal{M}_{m,t}$ to state $\mathcal{M}_{m,t+1}$ *conditional* on the action played a. This transition matrix is also indexed by time because CCPs and the transition processes of exogenous states (d_{mt}, h_{mt}) are indexed by time. Two transition matrices are necessary: for incumbents conditional on staying in the market, and for potential entrants conditional on entering. Denote the transition matrices for incumbents and entrants $\mathbf{F}^i_{m,t}$ and $\mathbf{F}^e_{m,t}$ respectively. If a firm plays a terminal action (that is, a potential entrant stays out) the continuation value is zero, therefore, knowledge of this transition matrix is not necessary.

3. For each market m and period t, solve for the ex-ante value function of an incumbent. For period $t \geq T$, the ex-ante value function solves the system of equations

$$\begin{aligned} \mathbf{V}^i_{m,T} &= \boldsymbol{\pi}_{m,T} - FC_m + \beta \mathbf{F}^i_{m,T} \mathbf{V}^i_{m,T} \\ &= \left(I - \beta \mathbf{F}^i_{m,T}\right)^{-1} [\boldsymbol{\pi}_{m,T} - FC_m], \end{aligned} \tag{15}$$

   where I is the identity matrix, $\boldsymbol{\pi}_{m,T}$ is a vector of variable profits in each state and FC_m is a market-specific fixed cost.[21] For period $t \leq T-1$, the ex-ante value function is obtained recursively as

$$\mathbf{V}^i_{m,t} = \boldsymbol{\pi}_{m,t} - FC_m + \beta \mathbf{F}^i_{m,t} \mathbf{V}^i_{m,t+1}. \tag{16}$$

4. Update the conditional choice-specific value function. Let $\mathbf{v}^e_{m,t}$ denote a vector collecting the choice-specific value function from entering in market m in period t. This vector satisfies the equality

$$\mathbf{v}^e_{m,t} = -EC_m + \beta \mathbf{F}^e_{m,t} \mathbf{V}^i_{m,t+1} \tag{17}$$

[21] We do not model exit decisions, therefore, we rule out firm-specific private information shocks for incumbents.

5. Update the vectors of CCPs as

$$\mathbf{P}'_{m,t} = \frac{\exp\left(\frac{\mathbf{v}^e_{m,t}}{\theta^\epsilon_m}\right)}{1+\exp\left(\frac{\mathbf{v}^e_{m,t}}{\theta^\epsilon_m}\right)} \tag{18}$$

If the maximum absolute difference between $\mathbf{P}_{m,t}$ and $\mathbf{P}'_{m,t}$, across periods $t = 1, \ldots, T$, is less than the pre-defined tolerance level (10^{-4}), the procedure stops and $\left(\mathbf{P}'_{m,t}\right)_{t=1,\ldots,T}$ is saved. If not, define updated CCPs as a convex combination of old and new CCPs $\eta\mathbf{P}_{m,t} + (1-\eta)\mathbf{P}'_{m,t}$ for each player i and return to Step 2.

As markets are independent, we can solve the model for each market separately. For our counterfactual analysis, we initialize this algorithm at a large number of starting values and iterate to a fixed point.

Estimation Approach. The objective of the dynamic game estimation is to recover the level of fixed costs FC_m and the scale parameter of the firm-specific shock θ^ϵ_m. We denote these parameters $\boldsymbol{\theta}_m \equiv (FC_m, \theta^\epsilon_m)$. Our baseline estimator is a pseudo-likelihood procedure following Aguirregabiria and Mira (2007). This procedure relies on a similar iterative procedure as the one used to solve the model, with an added estimation step.

Given a vector of CCPs and structural parameters in iteration k, denoted $(\mathbf{P}^{(k)}, \boldsymbol{\theta}^{(k)})$, we iterate over the following steps.

1. Update the transition matrix from state $\mathcal{M}_{m,t}$ to state $\mathcal{M}_{m,t+1}$ *conditional* on the action played a. As in the solution method, we store two transition matrices: for incumbents conditional on staying in the market, and for potential entrants conditional on entering. Denote the transition matrices for incumbents and entrants $\mathbf{F}^{i,(k+1)}_{m,t}$ and $\mathbf{F}^{e,(k+1)}_{m,t}$ respectively.

2. Solve for the ex-ante value function of an incumbent *gross* of the fixed cost, denoted $\widetilde{\mathbf{V}}^{i,(k+1)}_{m,t}$.[22]

$$\begin{aligned}\widetilde{\mathbf{V}}^{i,(k+1)}_{m,T} &= \boldsymbol{\pi}_{m,T} + \beta\mathbf{F}^i_{m,T}\widetilde{\mathbf{V}}^{i,(k+1)}_{m,T} \\ &= \left(I - \beta\mathbf{F}^{i,(k+1)}_{m,T}\right)^{-1}\boldsymbol{\pi}_{m,T},\end{aligned} \tag{19}$$

[22]We compute the value function gross of the fixed cost in order to be able to express the probability of entry as a function of the fixed cost and maximize the pseudo-likelihood. If we were to use Equations (15) and (16), then FC_m would be subsumed in the value function and would not appear as a argument in the pseudo-likelihood.

and

$$\widetilde{\mathbf{V}}^{i,(k+1)}_{m,t} = \boldsymbol{\pi}_{m,t} + \beta \mathbf{F}^{i}_{m,t} \widetilde{\mathbf{V}}^{i,(k+1)}_{m,t+1}. \tag{20}$$

3. Estimate the structural parameters $\boldsymbol{\theta}^{(k+1)}$ using a pseudo-likelihood estimator where the probabilities that a potential entrant enters in market m and period t is given by

$$\frac{\exp\left(\left\{-EC_m + \beta \mathbf{F}^{e,(k+1)}_{m,t} \widetilde{\mathbf{V}}^{i,(k+1)}_{m,t+1} - \frac{\beta}{1-\beta} FC_m\right\} \Big/ \theta^{\epsilon}_{m}\right)}{1 + \exp\left(\left\{-EC_m + \beta \mathbf{F}^{e,(k+1)}_{m,t} \widetilde{\mathbf{V}}^{i,(k+1)}_{m,t+1} - \frac{\beta}{1-\beta} FC_m\right\} \Big/ \theta^{\epsilon}_{m}\right)} \tag{21}$$

4. If the maximum absolute difference between $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\theta}^{(k+1)}$ and between $\mathbf{P}^{(k)}_{m,t}$ and $\mathbf{P}'_{m,t}$ (based on Equation (21) under $\boldsymbol{\theta}^{(k+1)}$) is less than the tolerance level, stop the procedure. If not, return to step 1 using $\mathbf{P}^{(k+1)}_{m,t} = \mathbf{P}'_{m,t}$ and $\boldsymbol{\theta}^{(k+1)}$.

Standard errors are calculated by bootstrap where markets are re-sampled (30 replications). We initialize this algorithm at several number of starting values for $(\mathbf{P}^{(0)}, \boldsymbol{\theta}^{(0)})$ and do not encounter any convergence issues. One advantage of the iterated procedure above (e.g., imposing equilibrium restrictions) is that it is robust with respect to the consistency of the initial estimates of CCPs.

Nonetheless, the NPL algorithm may fail to converge if the fixed point corresponding to the data generating process is unstable (Pesendorfer and Schmidt-Dengler (2010)). To address this concern, we implement several alternative estimators: two-step estimators (1-PML and 1-MD) and the spectral algorithm recently proposed by Aguirregabiria and Marcoux (2021). The latter estimator does not iterate on the best-response mapping to attain a fixed-point, rather, it solves for the root of a nonlinear system of equations by a quasi-Newton method. As a consequence, the spectral approach can find unstable fixed points that would be unattainable by the NPL algorithm.

Define $\phi(\mathbf{P}) \equiv \mathbf{P} - \Psi(\mathbf{P}, \hat{\theta}(\mathbf{P}))$, where $\hat{\theta}(\mathbf{P})$ is the pseudo-likelihood maximizer given input CCP vector $\mathbf{P}$ and $\Psi(.,.)$ is the best-response mapping as a function of input CCP and structural parameters. To find the solution(s) to $\phi(\mathbf{P}) = 0$, spectral approaches are particularly useful because they do not require computing the large-dimensional Jacobian $\nabla\phi(\mathbf{P})$ as would be required in Newton's method.[23] In practice, we replace the updating step above (step 4, where $\mathbf{P}^{(k+1)} = \Psi(\mathbf{P}^{(k)}, \hat{\theta}(\mathbf{P}^{(k)}))$) by

[23] Newton's method updates the CCP as

$$\mathbf{P}^{(k+1)}_{m,t} = \mathbf{P}^{(k)}_{m,t} - [\nabla\phi(\mathbf{P}_{m,t})]^{-1}\phi(\mathbf{P}^{(k)}_{m,t}).$$

$$\mathbf{P}_{m,t}^{(k+1)} = \mathbf{P}_{m,t}^{(k)} - \alpha_k \phi(\mathbf{P}_{m,t}^{(k)})$$

where α_k, the spectral step, equals

$$\alpha_k = \frac{\Delta\mathbf{P}^{(k)\prime}\Delta\mathbf{P}^{(k)}}{\Delta\mathbf{P}^{(k)\prime}\Delta\phi(\mathbf{P}^{(k)})}$$

and $\Delta\mathbf{P}^{(k)} = \mathbf{P}^{(k)} - \mathbf{P}^{(k-1)}$, $\Delta\phi(\mathbf{P}^{(k)}) = \phi(\mathbf{P}^{(k)}) - \phi(\mathbf{P}^{(k-1)})$. We set the initial value α_0 to one.[24]

6.3 Demand Model

In this section, we estimate demand-side objects that are not determined by the dynamic equilibrium: the demand for bandwidth and the transition of the exogenous demand state variables.

Bandwidth Demand. We estimate the demand for bandwidth via instrumental variables regression. We choose the fairly standard assumption of log-linearity, because it provides the best fit of the data. In particular, we assume that the demand for bandwidth in market m in period t (measured in Gbps) is a function of the bandwidth price P_{mt} (in \$US), the number of cables operating in the market n_{mt}, and a set of demand factors indexed by the variable d_{mt}. It takes the log-linear form

$$Q_{mt}(d_{mt}, P_{mt}, n_{mt}) = \exp\left(d_{mt} + \sum_n \alpha_n \mathbb{1}\{n_{mt} = n\}\right) P_{mt}^{\alpha_p} \tag{22}$$

To capture potential non-linear effects of the number of cables n_{mt}, we include it as a categorical variable.[25] The empirical analog of this demand curve is

$$\log(Q_{mt}) = \alpha_p \log(P_{mt}) + \sum_n \alpha_n \mathbb{1}\{n_{mt} = n\} + \alpha_2 \mathbf{X}_{mt} + \gamma_m + \gamma_{r(m)t} + \epsilon_{mt} \tag{23}$$

where $\mathbf{X}_{mt}$ are demand shifters, γ_m is a market-specific fixed effect, and $\gamma_{r(m)t}$ is a region-pair by year fixed effect. The term $\mathbf{X}_{mt}$ includes fixed broadband subscriptions, measures of Gross Domestic Product (GDP), aggregate trade flows, the number of data centers, where we restrict the coefficients on these variables to be the same for both countries forming the

[24]The spectral approach replaces the inverse of the Jacobian matrix $[\nabla\phi(\mathbf{P})]^{-1}$ by α_k. As shown in Aguirregabiria and Marcoux (2021), $1/\alpha_k$ approximates a Rayleigh quotient of $\nabla\phi(\mathbf{P})$: it is a weighted average of the eigenvalues of this matrix.

[25]The omitted category is "zero undersea cables." Including observations for market where undersea cables have not yet entered allows us to control for the baseline level of demand through other indirect paths.

market. It also include bilateral measures such as the distance between the two countries, whether they share a common language, and whether they share a land border (in specifications without market fixed effects). Region-pair by year fixed effects allow us to capture regional trends, e.g. changing composition of buyers (internet backbone providers, content providers, private enterprises) over time, or regional transient demand shocks.

We address the issue of endogeneity of prices via instrumental variables.[26] In particular, bandwidth prices are instrumented using marginal costs of electricity generation, a cost shifter. We use panels of electricity generation shares (by coal, gas, fuel) at the country-level, and time series data on prices of coal, gas, and oil, as detailed in Section 4.

Table 4 reports the results of the estimation of Equation (23). We show the OLS specification in columns (1) to (3). We find that demand shifters such as country-level broadband subscriptions, GDP, and the number of data centers have a positive effect on bandwidth demand.

The first and second stage of the IV regression are shown in columns (4) and (5). The first-stage suggests that electricity costs are strong predictors of bandwidth prices.[27] Once the endogeneity of bandwidth prices is accounted for, the price elasticity increases (in absolute value) to -1.36. This is consistent with the expected direction of the bias of the OLS regression. Column (5) also reports the first-stage F-statistic for the weak identification test which indicates that the instrument strongly predicts the endogenous variable.

With respect to the role of the number of cables, entry of additional cables has a positive and significant effect on demand holding bandwidth prices fixed. This corresponds to a "market expansion" effect as derived in Section 2. Figure 3 (right panel) shows the effect of the number of cables on bandwidth demand. We plot the coefficients from Table 4 along with those from a polynomial specification. The marginal effect of the number of undersea cables is decreasing with the number of cables indicating decreasing marginal returns from network diversity. The effect of adding a second cable (holding prices fixed) on bandwidth demand expands demand as much as a 23.5% decrease in bandwidth prices, adding a third cable is equivalent to a 20.4% decrease in prices, and adding an eighth cable is equivalent to an 6.1% decrease in prices.

We also perform the demand estimation on a subsample of markets, that we denote as "Top Routes."[28] These correspond to the top ten markets (by used bandwidth) connecting

[26]In our setting, sources of price endogeneity could be the correlation between demand and cost shocks. If marginal costs or the price elasticity depend on the quantity produced, demand shocks may also be transmitted to prices.

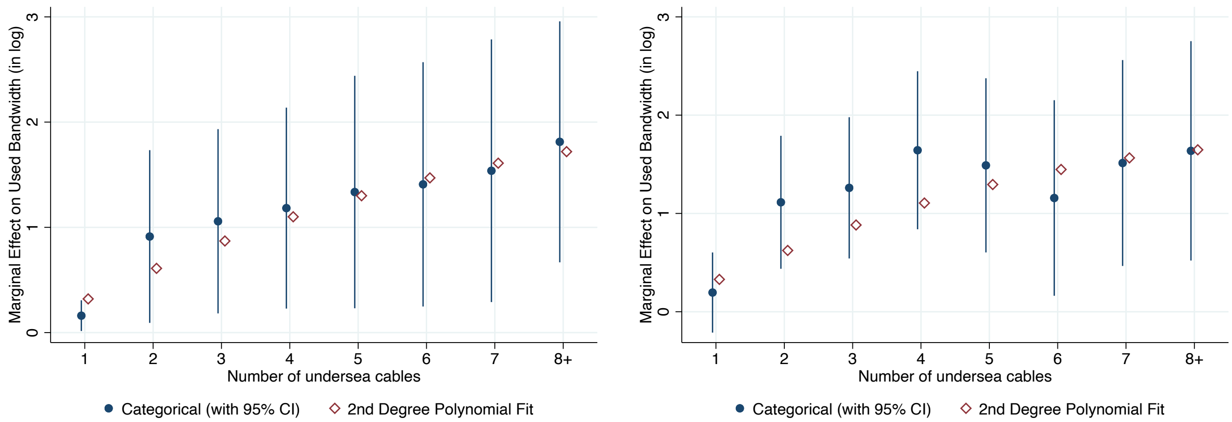
[27]We note that bandwidth prices are more correlated with the change in electricity costs. One interpretation is that lagged electricity costs are strongly positively correlated with current bandwidth prices.

[28]Appendix A provides detail about our market definition and gives examples of markets included in the final sample.

Table 4: Estimation Results: Used Bandwidth (in log) for All Routes

	OLS			IV	
	(1)	(2)	(3)	(4) First-stage	(5) Second-stage
Bandwidth Price (10G, log)	-0.926***	-0.733***	-0.214***		-1.360***
	(0.0876)	(0.0605)	(0.0530)		(0.295)
Number of undersea cables					
One cable	0.388	0.299	0.158	0.0165	0.195
	(0.669)	(0.802)	(0.167)	(0.0808)	(0.208)
Two cables	0.631	0.665	1.072**	0.0198	1.114**
	(0.666)	(0.816)	(0.338)	(0.110)	(0.345)
Three cables	1.112	1.072	1.412***	-0.135	1.261***
	(0.687)	(0.806)	(0.357)	(0.129)	(0.367)
Four cables	1.322	1.299	1.835***	-0.171	1.643***
	(0.819)	(0.853)	(0.407)	(0.155)	(0.410)
Five cables	1.413	1.386	1.765***	-0.243	1.489***
	(0.731)	(0.842)	(0.434)	(0.192)	(0.452)
Six cables	1.515*	1.416	1.561**	-0.363	1.157*
	(0.741)	(0.890)	(0.500)	(0.199)	(0.508)
Seven cables	2.566**	1.507	2.080***	-0.507*	1.513**
	(0.772)	(0.831)	(0.503)	(0.238)	(0.534)
Eight or more cables	1.460	1.504	2.305***	-0.589**	1.637**
	(0.835)	(0.947)	(0.506)	(0.227)	(0.569)
Demand factors					
Fixed Broadband Subscriptions (log)		0.370**	0.208	0.174*	0.411**
		(0.116)	(0.111)	(0.0771)	(0.133)
GDP (log)		-0.0646	0.932**	0.0221	0.972**
		(0.135)	(0.302)	(0.197)	(0.300)
Aggregate trade flow (log)		-0.00272	-0.0603	0.0261	-0.0317
		(0.0854)	(0.0496)	(0.0378)	(0.0770)
Number of data/cloud centers (log)		0.622***	0.165*	0.0194	0.193
		(0.0800)	(0.0800)	(0.0633)	(0.115)
Distance (km, log)		-0.0203			
		(0.134)			
Common official language		0.391			
		(0.214)			
Contiguous		0.694**			
		(0.264)			
Electricity price (log)				0.0510*	
				(0.0210)	
% change in Electricity price				-0.218***	
				(0.0349)	
Country Pair FEs	No	No	Yes	Yes	Yes
Region Pair × Year FEs	No	No	Yes	Yes	Yes
Weak Identification test					22.31
Endogeneity test (p-value)					26.8 (0)
R^2	0.39	0.71	0.97		0.96
Adjusted R^2	0.39	0.71	0.97		0.95
Observations	4908	3863	3849	3863	3863

Note: The unit of observation is a country pair by quarter. Standard errors are clustered at the country pair level. Distance corresponds to the bilateral distances between countries, calculated as a weighted arithmetic average of the geodesic distances between the main cities in these countries, where population weights are used. For unilateral variables (GDP, fixed broadband subscriptions, data centers, electricity prices), we restrict the coefficient to be the same for both countries in a pair.



(a) Top routes

(b) All routes

Figure 3: The marginal effect of the number of operating undersea cables on used bandwidth (in log). The specifications with cable count and a (2nd degree) polynomial fit are shown.

to each region in the world. This subsample accounts for a large share of global bandwidth usage and allows us to verify that our results are not driven by outlier markets. The results of the demand estimation, shown in Table A7 of Appendix C and on the left panel of Figure 3, are quantitatively similar for this subsample of markets.

Finally, we conduct robustness checks with respect to effect of the number of cables. As explained in Section 6.1, our baseline specification assumes that firms make entry decisions before the realization of current transient demand shocks: the exact timing of entry, therefore, reflects idiosyncratic shocks (e.g., time-to-build, delays in construction) that are unrelated to unobserved demand shocks. Nonetheless, we can address concerns of endogeneity of the number of cables via instrumental variables. We use the occurrence of cable faults as an exogenous shifter of entry and subsequent market structure: we find support that markets hit by cable disruptions see significantly less entry in the next period, one reason being that disruptions tie up cable ships in repairs instead of construction of new cables. Table A8 in Appendix C shows the results, where new cable entry is instrumented by lagged cable faults.[29] We find that the effect of the number of cables (column (6)) remains positive and consistent with our baseline estimate (column (3)).

Demand State variable. The demand state variable d_{mt} is defined as the intercept of the (log) demand curve

$$d_{mt} \equiv \widehat{\alpha}_2 \mathbf{X}_{mt} + \widehat{\gamma}_m + \widehat{\gamma}_{r(m)t} + \widehat{\epsilon}_{mt} \tag{24}$$

[29] Due to the limited scope of the cable faults data, the sample size is smaller. We also include the number of cables as a continuous variable in the IV regression.

The residual of the IV regression is included in d_{mt} as it captures omitted time-varying demand shifters. The market and region-year dummies are also included to control for market-specific unobserved heterogeneity and regional trends. Using the demand estimation toward the construction of a demand state variable has the benefit of allowing us to combine several observed and unobserved demand factors affecting bandwidth demand (i.e., the $\mathbf{X}_{mt}$'s and fixed effects) into a single index. This is particularly helpful in anticipation of estimation and computation of equilibria of the dynamic game as it drastically reduces the size of the state space.

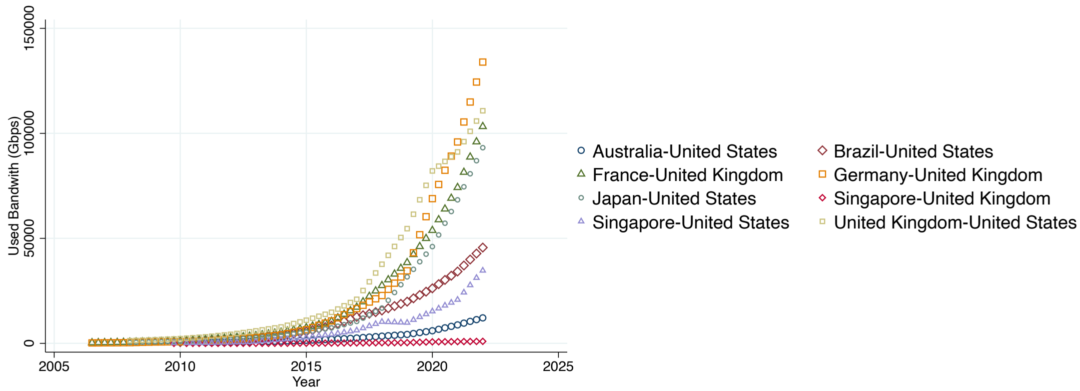
Figure 4 shows bandwidth demand Q_{mt} in level and logarithm (top and middle panels) and the demand state d_{mt} (bottom panel) for a sample of markets. Demand for bandwidth increases exponentially over the sample period, with some degree of heterogeneity across markets of different sizes. This exponential growth is the result of both the widespread adoption of the internet on the demand side, with increased usage due to new activities such as social media, streaming, and cloud computing and storage; and on the supply side, advancements in telecommunication technologies which enables faster and more cost-efficient data transmission. The demand state d_{mt} shown in the bottom panel corresponds to the portion of total bandwidth demand Q_{mt} that is not explained by falling prices (due to cost improvements) or greater network diversity but instead is driven by the evolution of exogenous demand factors.

Transitions of the demand state. We specify the evolution of the demand state variable d_{mt} as a first-order autoregressive process (AR(1)), where the auto-regressive parameter is homogeneous across markets but the mean varies over market (i.e., market-specific drift term), as follows

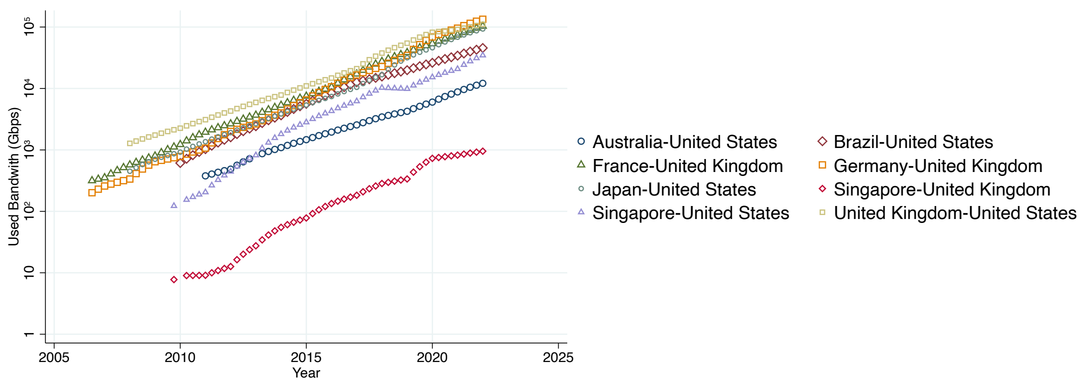
$$d_{mt} = \alpha_m + \rho_d d_{m,t-1} + \nu_{mt}. \tag{25}$$

The parameters of this AR(1) process are estimated by full maximum likelihood.[30] The estimate of the autoregressive coefficient is 0.933 (se = 0.005). Other estimation methods (within group, first-difference, and OLS) give similar results. We also explored alternative specifications with homogenous drifts or with period-specific fixed effects, as well as the inclusion of a time trend. These OLS specifications are shown in Table 5. The time trend is not statistically significant and the addition of time fixed effects does not improve the fit of the model. In the remainder of the estimation, we proceed using the specification in column (2) of Table 5. Section 6.6 shows the model fit and provides evidence that this specification fits the data well.

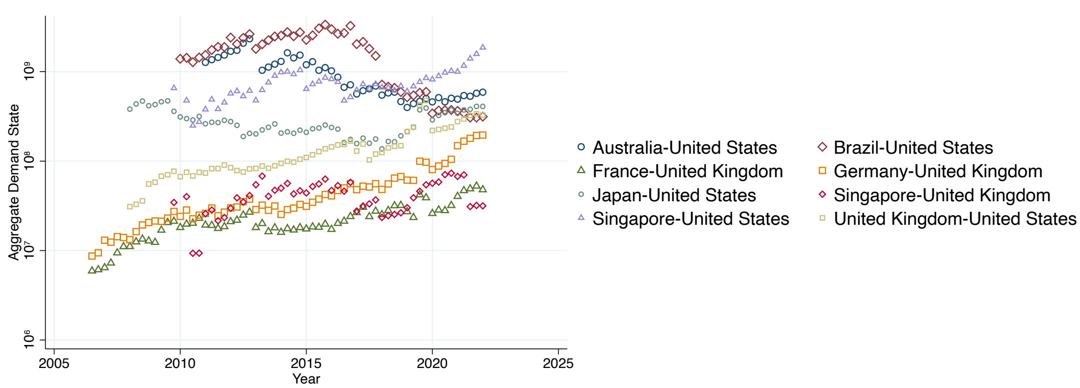
[30]We have a long panel (16 years of quarterly data) which alleviates concerns related to the incidental parameter problem.



(a) Used bandwidth Q_{mt} in level



(b) Used bandwidth Q_{mt} in log scale



(c) Demand state d_{mt} (Gbps, log scale)

Figure 4: Used bandwidth and demand state over time for a sample of markets

Table 5: Transition process for the demand state d_{mt}

	(1)	(2)	(3)	(4)
Demand state in $t-1$	0.989	0.933	0.942	0.932
	(0.00274)	(0.00724)	(0.00742)	(0.00809)
Time trend				0.000841
				(0.00208)
Market-level Intercept	No	Yes	Yes	Yes
Year-Quarter FEs	No	No	Yes	No
R^2	0.98	0.98	0.99	0.98
Adjusted R^2	0.98	0.98	0.99	0.98
Observations	4646	4644	4644	4644

In anticipation of the estimation of the dynamic game, we use the method in Tauchen (1986) to discretize this AR(1) process and obtain the transition matrix of the discretized variable on a finite support.[31] The support of the demand state is allowed to vary by market.

6.4 Marginal Costs Estimates

In this section, we present the estimation of cable operators' marginal costs of bandwidth. To estimate these costs, we combine the demand estimates from the previous section and the first-order conditions of the firms' period profit maximization. We assume Cournot competition between cable operators that are active in period t. This conduct assumption is motivated by features of the industry and discussion with industry experts: cable operators light capacity in advance and every period (i.e., either activate new fiber pairs or new wavelengths on lit fiber pairs) and are committed to selling their lit capacity. This is more in line with strategic behavior under Cournot competition. We provide tests of this conduct model in Section 6.6.

Marginal costs of supplying bandwidth are assumed to be symmetric across suppliers but can vary by market and time period to capture heterogeneity in cost factors across regions and technological advances as discussed below. The symmetry assumption is motivated by our data source which does not report cable-level market shares for confidentiality reasons. We relax the cost symmetry assumption and conduct robustness checks in Appendix B.3.

[31] We eliminate the market-specific drift term by rewriting Equation (25) as

$$d_{mt} - \frac{\alpha_m}{1-\rho_d} = \rho_d \left(d_{m,t-1} - \frac{\alpha_m}{1-\rho_d} \right) + \nu_{mt},$$

and estimating the transition matrix for the de-meaned process $\widetilde{d}_{mt} = d_{mt} - \frac{\alpha_m}{1-\rho_d}$.

Recovering Marginal Costs. In market m and period t, firm i's first-order condition with respect to its output q_{imt} is

$$p_{mt}(q_{imt}, q_{-i,mt}) + \frac{\partial p_{mt}}{\partial q_{imt}} q_{imt} = mc_{mt}. \tag{26}$$

where, under a symmetric equilibrium, q_{imt} equals Q_{mt}/n_{mt}. Equation (26) is used to infer the marginal costs mc_{mt} of suppliers active in each market and time period. Figure 5 shows box-plots of marginal costs estimates for a subsample of region pairs over time.

The downward trend in marginal costs can be ascribed to two main factors. On the one hand, economies of scale may allow cable operators to expand capacity with decreasing unit costs. On the other hand, technological advances also reduce the cost of deploying additional bandwidth: these advances include various fiber-optic transmission system upgrades as well as improvements in software-defined networking.[32] We also note important heterogeneity in costs across region pairs with the lowest costs attained by markets in the Trans-Atlantic and Intra-Europe regions and some of the highest costs of bandwidth attained by markets in the Intra-Asia-Oceania and Latin America-North America regions. Within a region pair, we also note that costs can vary by an order of magnitude across markets.

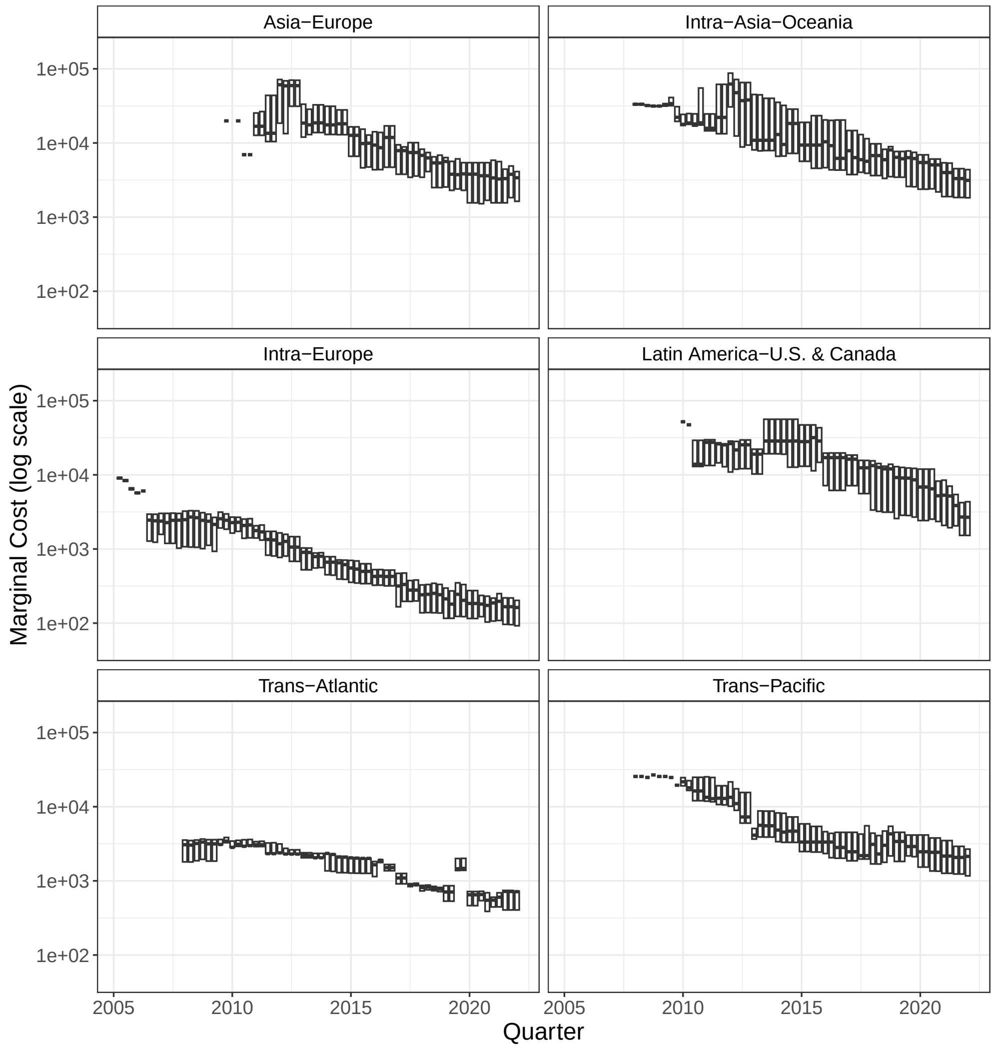
Marginal Costs Function. We investigate how marginal costs vary with cable characteristics and other cost shifters. The following specification for the marginal cost function is used

$$\log(mc_{mt}) = \gamma_0 \mathbf{W}_{mt} + \gamma_q \log(q_{mt}) + \eta_m + \eta_{r(m)t} + \omega_{mt}, \tag{27}$$

where $\mathbf{W}_{mt}$ are exogenous cost shifters, q_{mt} is the quantity supplied by a given firm, η_m and $\eta_{r(m)t}$ are market fixed effects and region by time fixed effects, and ω_{mt} are unobserved cost shocks at the market-period level. This specification allows marginal costs to depend on quantity to capture potential economies of scale. We estimate this specification by OLS and show the results in Table 6. As expected, we find a positive correlation between marginal

[32] Fiber-optic transmission system upgrades are technologies and methods used to increase the capacity and improve the performance of *existing* fiber-optic transmission systems. Examples include Wavelength division multiplexing, a technique that allows multiple wavelengths to be transmitted over a single fiber-optic cable at the same time; Optical amplifiers, devices that boost the strength of the light signals that are transmitted over a fiber-optic cable; Polarization-multiplexing, a technique used to transmit two different data streams over a single fiber using different orientations of light waves (or polarization); Space-division multiplexing, a technology used to increase the capacity of the cable by using multiple cores within a single optical fiber. These techniques are generally considered to be relatively low-cost as they do not involve significant capital investments. Software-defined networking can be used to manage the traffic on the cable by providing a centralized control plane that can make decisions about how to route traffic based on real-time information about the network's state. This allows for greater flexibility in managing the network and can enable more efficient use of network resources.

Figure 5: Marginal cost estimates by region pair



Notes: The figures display box-plots of marginal costs estimates for a subset of region pairs over time. Each white bar shows the first and third quartiles of the marginal costs distribution across markets (country pairs) in a given region pair. The black segment shows the median of this distribution. The unit is US$ per quarter at 10Gbps capacity.

costs and electricity prices, cable length, and the number of (contemporaneous) cable faults. In specifications (5) and (6), marginal costs are also decreasing in the amount of bandwidth supplied by a given cable, suggesting the presence of economies of scale. However, this effect is not robust to the inclusion of market-level fixed effects (specification (7)).[33]

Transition of Marginal Cost State. In anticipation of the estimation of the dynamic game, we recover the transition process of marginal costs. We define the cost state variable h_{mt} as the logarithm of the marginal cost, that is,

[33]Because q_{mt} may be correlated with the unobserved cost shock, we also run specification (7) instrumenting q_{mt} with demand-side shifters. The results remain quantitatively similar.

Table 6: Determinants of marginal costs

	Dependent variable: Marginal cost						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Used bandwidth	-0.384				-0.368	-0.124	0.006
	(0.008)				(0.008)	(0.006)	(0.015)
Electricity Price		0.530			0.248	0.072	0.193
		(0.025)			(0.022)	(0.033)	(0.037)
% Change in Electricity Price		-0.211			-0.144	-0.145	-0.708
		(0.073)			(0.060)	(0.105)	(0.096)
Cable length			0.585		0.538	-0.047	
			(0.017)		(0.016)	(0.018)	
Number of cable faults				0.140	-0.007	0.009	0.010
				(0.034)	(0.025)	(0.015)	(0.013)
Region Pair FE	No	No	No	No	No	Yes	No
Time FE	No	No	No	No	No	Yes	No
Region Pair × Time FE	No	No	No	No	No	No	Yes
Market FE	No	No	No	No	No	No	Yes
R^2	0.333	0.163	0.190	0.004	0.586	0.837	0.975
Adjusted R^2	0.333	0.163	0.190	0.004	0.585	0.834	0.968
Within R^2						0.087	0.028
Observations	4,777	4,493	4,777	3,714	3,430	3,430	3,430

Note: The unit of observation is the market (country pair) by quarter. All variables are in log (except change in electricity prices and number of cable faults). The number of cable faults is only available starting in 2013. Used bandwidth corresponds to the amount supplied by a single firm. Cable length is averaged over all cables operating in a given market.

$$h_{mt} \equiv \log(mc_{mt}) \tag{28}$$

The evolution of the cost state variable h_{mt} is specified as a non-stationary AR(1) process with a time trend, where the auto-regressive parameter is homogeneous across markets but the mean varies over market (market-specific drift term) as follows,

$$h_{mt} = \tau_m + \rho_h h_{m,t-1} + \delta t + \xi_{mt}. \tag{29}$$

The parameters of the AR(1) process are estimated by maximum likelihood. Other estimation approaches (within group, OLS) give similar results. The autoregressive coefficient estimate is 0.890 (se = 0.001) and the time trend δ estimate is -0.023 (se $= 6 \cdot 10^{-6}$). As with the demand state variable d_{mt}, we discretize the AR(1) process for h_{mt} to allow for tractable estimation of the dynamic game. Because the process is non-stationary, we obtain the transition matrix for each period. The support of the cost state h_{mt} is allowed to vary by market. Section 6.6 shows the model fit and provide evidence that this specification fits the data well.

6.5 Dynamic Investment Cost Estimates

In the last step of our estimation approach, we combine the static estimates (from the demand model and firms' marginal costs) to construct period profits and incorporate these profits into firms' dynamic game of entry. The dynamic parameters of interest are the entry costs, fixed costs, and the scale of the private information shocks.

As discussed in the identification of the model (Section 6.1), we cannot separately identify entry costs from fixed costs because we do not observe sufficiently many exit events. Therefore, we start by recovering entry costs directly from data on cable construction costs. Next, we follow the methodology outlined in section 6.2 to estimate the level of fixed costs and the scale of the logit shocks.

Entry costs. Our data source contains information on cable construction costs for 47% of active cables. As described in Section 3, the major cost drivers in constructing an undersea cable system can be divided into: (1) the "wet plant", undersea components including fiber, cable, repeaters, and branching units, (2) the "dry plant", which include Submarine Line Termination Equipment, power feeding equipment, monitoring systems, and network protection equipment, and (3) marine operations, tasks related to planning and installation of the cable, such as route surveying and cable loading, laying, and burial. The wet plant is the most distance-sensitive and usually the most costly component of a subsea cable construction project. This directly determines the amount of cable, fiber, and repeaters deployed. Longer spans require more cable and more repeaters. Figure 6 shows cable-level construction costs as a function of cable length (in level and log). The mean cable construction cost is $244 million and the median construction cost is $66 million.

We use the available construction costs data to predict the entry costs EC_m for all remaining markets analyzed. This is operationalized by regressing construction costs on cable length and region fixed effects. The latter are meant to capture differences in topology across regions. Indeed, the geographical features of a cable route play an important role in calculating the cost of fiber installation. For instance, in shallower waters and in fault-prone areas the cable must be armored and buried in trenches for protection. To predict entry costs, we use the average cable length over all operating cables in a given market, as the cable length data is available for all active cables.[34]

Fixed costs and scale parameter. Fixed operational costs include administrative costs for

[34] We interpret the cable constructions costs data as representative of the "list" price of constructing the cable. These data are, therefore, net of the firm-specific private information shock.

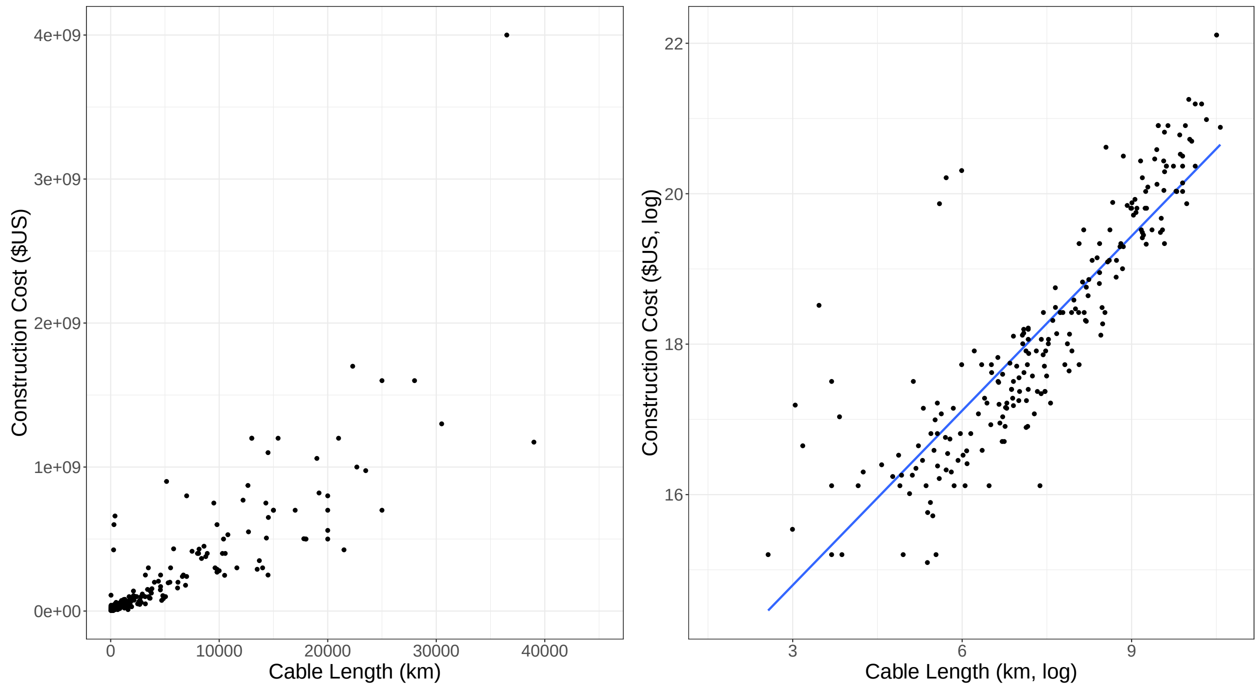


Figure 6: Cable construction costs and cable length in level (left) and logarithm (right).

office equipment, administrative and network staffing, marketing, legal and regulatory fees,[35] as well as network operation costs for maintenance and upgrade of undersea equipment, landing stations, and network operations facilities (see footnote 32 for examples of system upgrades).

Given the large heterogeneity in cable system sizes (e.g., Belgium-United Kingdom and Japan-United States), we allow the fixed costs and the scale of the logit shocks to vary across market. In particular, we parameterize the (quarterly) fixed costs and scale parameter as linear functions of the entry costs: $FC_m = \delta_{FC} EC_m$ and $\theta_m^\epsilon = \delta_\epsilon EC_m$.

Although the parameters $(\delta_{FC}, \delta_\epsilon)$ are separately identified because static profits are "observed" when estimating the dynamic game (see Equation (21)), in practice, we encountered the presence of multiple local maxima of the likelihood function in the ratio $\frac{\delta_{FC}}{\delta_\epsilon}$.[36] To address this problem, we estimate the scale parameter δ_ϵ under different values of δ_{FC} which are informed by discussion with industry professionals and various cable-level cost reports. In particular, we choose upper and lower bounds for δ_{FC} so that our cost estimates are in line with typical industry ratios of operating expenses and capital expenses as discussed in detail in the next section.

[35] The Federal Communications Commission fees currently run a little more than $53,000 for systems that have between 250 Gbps and 1.5 Tbps of lit capacity.

[36] We conjecture that this due to the entry costs and fixed costs (term $-EC_m - \frac{\beta}{1-\beta} FC_m$) dominating the remaining term (continuation value) that pins down δ_ϵ.

Given a value for δ_{FC}, the scale parameter is estimated precisely and the iterations of our estimation algorithm converge to the same estimate in a short number of iterations. Estimates of the scale parameter are shown in Table 7 for three different estimators.[37] In all bootstrap iterations, the NPL and spectral algorithms converge to the same fixed point of the NPL mapping (regardless of starting values) and, therefore, the corresponding pseudo-likelihood maximizer is virtually identical (up to the fifth digit). Under the assumption that $\delta_{FC} = 0.2\%$, annual fixed costs are in the range of \$1 million (the mean is \$1.36 million and the median is \$1.1 million) in line with estimates reported in TeleGeography (2022). We proceed, in the counterfactual analysis, by using estimates obtained with the NPL algorithm under $\delta_{FC} = 0.2\%$.

Table 7: Parameter estimates from the dynamic model

	$\delta_{FC} = 0$	$\delta_{FC} = 0.2\%$	$\delta_{FC} = 0.8\%$
Scale of entry shocks: $\theta^{\epsilon}_m = \delta_\epsilon EC_m$			
Parameter δ_ϵ			
1-PML	0.226 (0.010)	0.235 (0.010)	0.262 (0.011)
NPL Algorithm	0.225 (0.010)	0.234 (0.010)	0.261 (0.011)
Spectral Algorithm	0.225 (0.010)	0.234 (0.010)	0.261 (0.011)

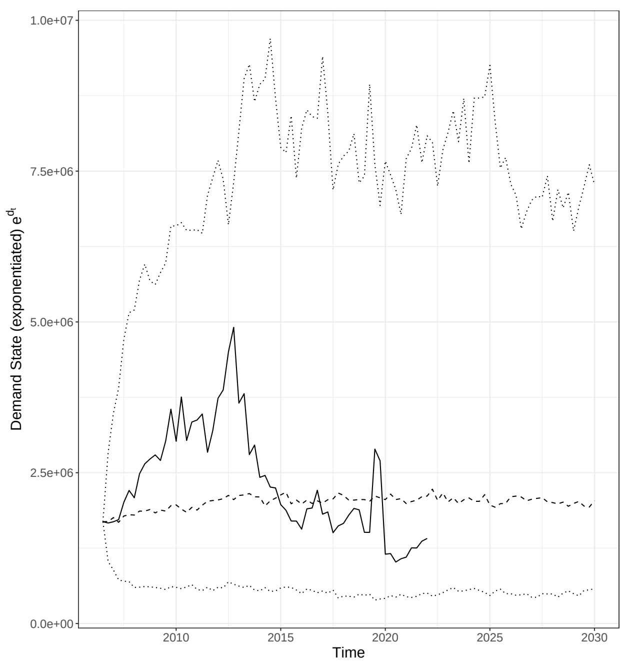
Note: Standard errors, in parenthesis, are obtained by bootstrap (30 replications).

6.6 Model Fit and Assumptions

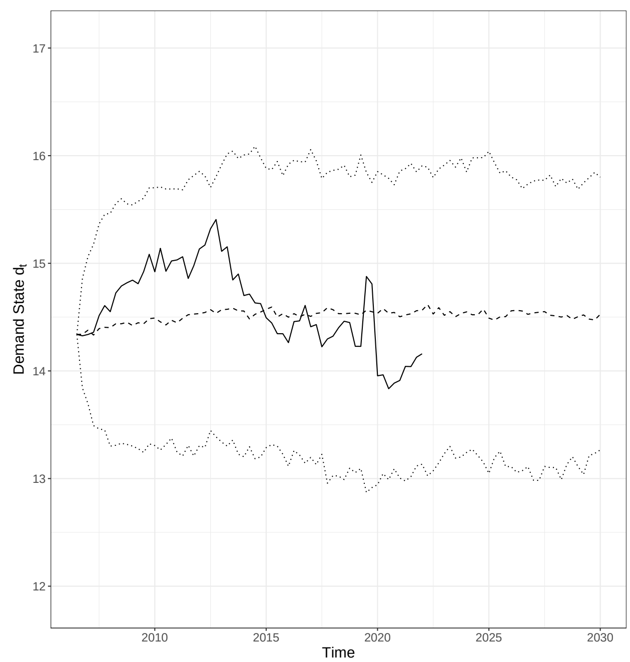
In this section, we examine the fit of the model and discuss our modelling assumptions. First, we consider the estimated transition processes of the demand and cost states (d_{mt}, h_{mt}). These constitute important inputs into the dynamic game. We use the fitted AR(1) processes to simulate these two state variables forward and compare simulation results to the realization of the states in the data. Figure 7 shows the results for the market "Belgium-United Kingdom." The top two figures show the demand state d_{mt} (exponentiated in Gbps, and in level), and the bottom figures show the cost state h_{mt} (exponentiated in US\$, and in level) in the data, along with 95% confidence intervals from our simulations (dotted lines) and the median simulated value (dashed line).[38] We show similar simulations exercises for a subsample of markets in Figures A5 and A6 of Appendix C. Overall, the estimated transition processes captures the time series variation and the heterogeneity across markets well.

[37] Aside from the 1-PML, we also experiment with a minimum distance estimator (Pesendorfer and Schmidt-Dengler (2008)). The latter gives quantitatively similar results.

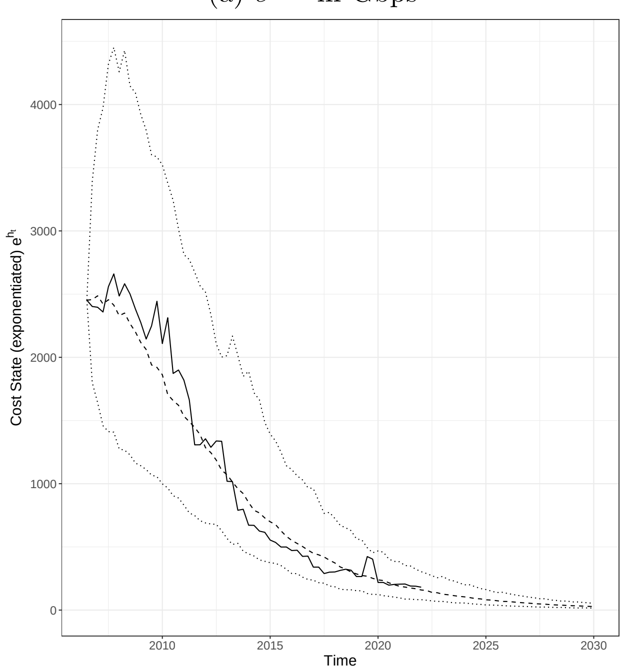
[38] The last period in the data is 2021-Q4. For the simulations, we iterated forward until 2030.



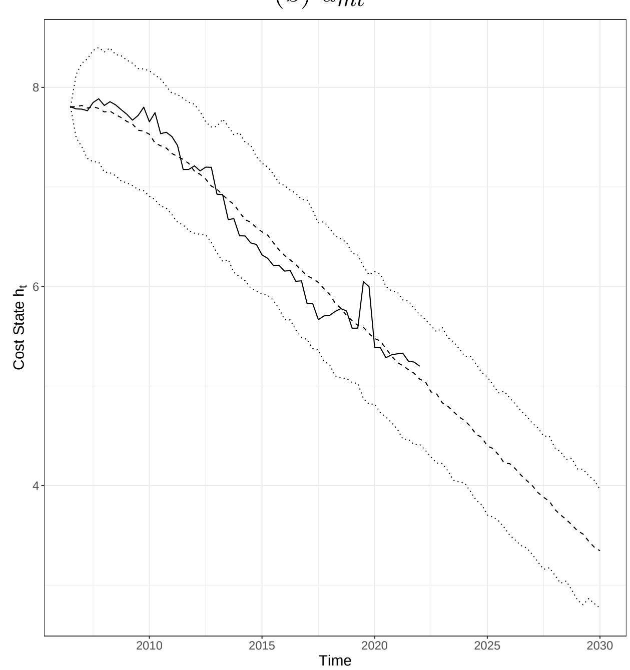
(a) $e^{d_{mt}}$ in Gbps



(b) d_{mt}



(c) $e^{h_{mt}}$ in US$



(d) h_{mt}

Figure 7: Demand State and Cost State over time for the market Belgium-United Kingdom
Notes: The figures show the demand state d_{mt} and cost state h_{mt} over time (solid lines) in level and exponentiated. The dotted lines give the 95% confidence interval from simulations using the transition processes. The dashed line shows the median simulated value.

Second, we compare the magnitude of the various cost estimates we obtain to estimates from industry reports and cable financial investment plans (Reverdy and Skenderoski (2015), Seixas (2015)). These reports typically divide costs into capital expenditures (CAPEX) which are payable to the main manufacturer and cable installation supplier, and operational expenditures (OPEX) which are incurred *annually* across the cable's lifespan. Estimates of OPEX to CAPEX ratios from these reports range from 4% to 7% depending on the system size and characteristics. For comparison purposes, we compute the ratio of per-year costs (variable costs and fixed costs) to the entry costs. Under the assumption $\delta_{FC} = 0.2\%$, we find an average ratio of 8% and median ratio of 3%. Under the assumption $\delta_{FC} = 0.8\%$, the mean of per-period costs to entry costs ratio is 10% and the median ratio is 5%.[39] Overall, the magnitude of our costs estimates is of the same order as that found in industry reports.

Third, given our structural parameter estimates, we solve for the model equilibrium and compare the model predictions to the actual data. This is a useful way of evaluating the assumptions embedded in the model, in particular, those concerning information, timing, functional forms, etc. The equilibrium CCP (solved for) are used to simulate the industry forward from the initial industry state and predict the expected number of cables in the last period of the sample. Figure 8 shows the number of cables in the last period in the data against the (simulated) expected number of cables predicted by the equilibrium of the dynamic game. We find that the model performs well, predicting slightly too much entry into markets that have only one cable in the data, but matching the number and distribution of cables quite closely.

Fourth, we test the conduct assumption (symmetric Cournot) using the approach introduced in Pakes (2017).[40] We leverage the dual facts that one can obtain an independent estimate of the markup from the demand parameters, and that our conduct assumption implies, via the first-order condition, that this markup has a coefficient of one in the quantity-setting equation. Rewriting firms' first-order condition (Equation (26))

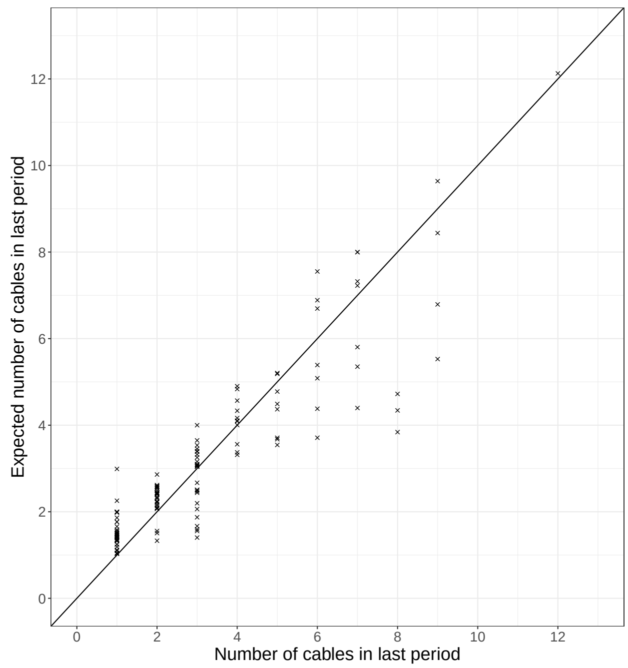
$$p_{mt} = mc_{mt} - \frac{\partial p_{mt}}{\partial q_{imt}} q_{imt} = \gamma_0 \mathbf{W}_{mt} + \omega_{mt} - \frac{\partial p_{mt}}{\partial q_{imt}} q_{imt} \tag{30}$$

where $\mathbf{W}_{mt}$ are exogenous cost factors (as in Table 6), ω_{mt} are unobserved cost shocks, and $-\frac{\partial p_{mt}}{\partial q_{imt}} q_{imt}$ is the markup. The latter variable can be calculated directly using the demand

[39]We note the presence of a few outlier markets where the ratio is above one. Indeed, markets where the two countries are very close geographically, e.g., "Denmark-Sweden" or "Indonesia-Singapore" have relatively low entry costs due to short cable lengths, but may have high variable costs because substantial levels of bandwidth traffic is carried by these short cables.

[40]See also Villas-Boas (2007) for a similar approach. Due to limitation in the data, we cannot use more involved approaches (e.g., non-nested testing, model selection): in particular, without cable-level market shares, it is difficult to obtain price elasticities under a differentiated Bertrand assumption.

Figure 8: Model fit for the number of cables in the last period



estimates. Note that this variable is endogenous as it is a function of both unobserved demand and cost shocks. The test is implemented as follows: (1) use the demand estimates to construct the markup term $-\frac{\partial p_{mt}}{\partial q_{imt}} q_{imt}$, (2) regress prices on observed cost factors and the markup, where the markup is instrumented using exogenous demand shifters (in practice we use GDP, cloud/data centers, and broadband subscriptions), (3) test statistically whether the estimated coefficient for the markup term is equal to one. The regression table is included in Table A9 of Appendix C. The fit of the regression is high ($R^2 = 0.76$). The markup coefficient is 1.28 (se = 0.23), and we cannot reject the null hypothesis of equality to one (the $p-$value is 0.21). While this result does not constitute conclusive evidence in favor of our conduct assumption, it is reassuring that the data is not at odds with the Nash in quantites assumption.

Finally, we discuss the assumption of independent markets. In the industry model, cables in each market are assumed to play a separate Markov Perfect equilibrium and are not affected by the evolution of neighboring markets. A first concern regarding this assumption is that demand might be correlated across neighboring markets. This concerns is arguably alleviated by the control for market-level unobserved heterogeneity and regional-time trends in our demand estimation (Equation (23)). We allow for (anticipated) macro-economic shocks at the region-pair level that shift demand simultaneously for neighboring markets (e.g., U.S.-U.K., U.S.-France, U.S.-Ireland).

A second concern is that the demand for a focal market might be in part carried via indirect paths through a third country.[41] Since we observe only aggregate bandwidth flows between country pairs and not the exact path followed, we cannot directly verify this assumption in the data. Moreover, industry experts confirm that data on the exact routing of bandwidth would be extremely difficult if not impossible to compile. Nonetheless, we can provide suggestive evidence in support of this assumption. First, we examine the correlation between bandwidth growth in a focal market, on the one hand, and the number of operating cables in the focal market and neighboring markets, on the other hand. We expect the number of cables in the focal market to be positively correlated with bandwidth growth. By contrast, under our independence assumption, the number of cables in neighboring markets should not be correlated with bandwidth growth in the focal market. We find that this is indeed the case.[42]

Second, we compare the total potential capacity (e.g., the theoretical design capacity at maximum utilization which is available for a subset of cables in 2021) installed in a focal market to the used bandwidth in that market. Consistent with the aggregate figures shown in the last column of Table 1, potential capacity is on average four times larger than bandwidth demand per market, indicating that direct paths are not capacity-constrained in serving market demand.[43]

7 Counterfactual Analysis

In this section, we use the estimated model to examine the role of supplier diversification in shaping the dynamics of the industry and the amount of surplus generated. Specifically, we conduct two counterfactual exercises. The first one evaluates the importance of supplier diversification for investment in new undersea cable systems and consumer and producer surplus created between 2005 and 2021. The second one compares the level of entry in the market outcome to the socially optimal level of entry, and assesses whether market forces

[41] This issue is also typically present in economic studies of the airline industry when it comes to defining product markets. Williams (2022), for instance, filters markets to select homogeneous itineraries, where at least 75% of trips are non-connecting, nonstop trips. Non-connecting trips are trips where either the origin or destination is a connection.

[42] To implement this test in practice, we find, for each focal market, the largest country to which the pair is connected, which we refer to as "neighboring" country. "Largest" is defined by the total amount of bandwidth connected to the country in 2021. For example, for the focal market U.S.-France, the neighboring country would be the UK. Next, we regress bandwidth growth in the focal market on (1) the number of cables in the focal market and (2) the number of cables connecting the countries in the focal market to their neighboring country. We include market fixed effects to focus on within-market correlations between market structure and demand.

[43] We also note that transiting through a third party country discontinuously raises costs because IP transit costs (akin to tolls) need to be paid to the provider transmitting the data via the third country. This makes the use of a direct path (via an undersea cable), if available, more cost efficient.

provide enough incentives for network diversity. In this second exercise, we quantify the relative sizes of business-stealing and diversity externalities as sources of distortions in entry.

7.1 The role of network diversity

In the first counterfactual exercise, we assess how quantitatively important supplier diversification is in shaping the dynamic evolution of the industry and the amount of surplus created.

To implement this counterfactual exercise, we assume buyers cannot diversify their supplier base. That is, in the demand system (Equation (22)), we fix n_{mt} to be one regardless of the number of cables operating. Entry of new cable systems intensifies competition (i.e., raises output and reduce prices) but does not benefit buyers through increased supplier diversification. Therefore, the market expansion effect acting through supplier diversification is shut down. In the context of the model of Section 2, this is equivalent to assuming that the default risk is perfectly correlated across all cables in a given market. Given this counterfactual demand model, we solve for the nonstationary MPE of the game following the methodology outlined in Section 6.2.

Table 8 shows the outcomes under the equilibrium played in the data (DGP) and the counterfactual equilibrium (CF) without diversification. When buyers cannot diversify, entry does not expand demand as much as in the market outcome. We find that consumer surplus under the counterfactual scenario amounts to 66% of consumer surplus under the DGP: in other words, supplier diversification accounts for 34% of consumer surplus created over the sample period. Counterfactual entry probabilities are 16% lower (from a baseline mean entry probability of 4% in the DGP) because entry no longer expands demand via increased diversity.[44] Producer surplus does not vary significantly, however. This is due to the fact that, although variable profits are lower under the counterfactual equilibrium, entry costs and the private information shocks tend to dominate the continuation value from being an incumbent. Moreover, the reduction in entry rates is associated with increased market power which counters the drop in profits due to lower demand.

Average total surplus per market over the sample period (in discounted terms) is equal to 895 million US$. Preferences for supplier diversification accounts for 13% of total surplus created in the industry over the sample period.

[44] We focus here on the likelihood of entry rather than the change in the number of cables. The latter may also reflect the fact that we can simulate forward only starting from a date where price and quantity data are available (which in many instances occurs after 2005), see Appendix A.

Table 8: Counterfactual outcomes under no supplier diversification

	DGP		CF		CF/DGP	
	Mean	Median	Mean	Median	Mean	Median
Entry rates						
Number of cables in 2005-Q1 (data)	1.56	1.00				
Number of cables in 2021-Q4 (data)	3.20	3.00				
Expected number of cables in 2021-Q4	3.17	2.50	3.07	2.45	0.97	0.99
Expected probability of entry	0.033	0.014	0.020	0.013	0.84	0.93
Welfare (2005-Q1 to 2021-Q4) in millions of US$						
Consumer Surplus	508.87	59.77	219.45	41.37	0.66	0.69
Producer Surplus	385.63	279.82	385.79	287.94	1.00	1.00
Total Surplus	894.51	385.23	605.24	367.35	0.87	0.94
Consumer Surplus / Total Surplus	0.30	0.22	0.23	0.15	0.76	0.77

Note: The first two columns correspond to the data generating process (DGP) and show outcomes under the equilibrium played in the data. The following two columns (CF) show counterfactual outcomes when diversification is shut down. The last two columns show the ratio of the counterfactual outcomes to DGP outcomes. To compute the "expected probability of entry," the expected probability of entry in each period is computed given equilibrium state transitions and choice probabilities, then the average over all periods is calculated.

7.2 The efficiency of entry

In the second counterfactual exercise, we compare the equilibrium level of entry in the market outcome to the socially optimum level of entry. The objective is to evaluate whether market forces provide entrants with insufficient or excessive entry motives and compare the resulting level of supplier diversity to the one chosen by a social planner.

There are two sources of distortions. First, entry may be excessive due to standard business-stealing motives: i.e., when entry reduces incumbents' output, the private benefit exceeds the social benefit of entry. Second, entry may be insufficient due to a diversity effect: i.e., the marginal entrant creates surplus by increasing supplier diversity that they do not (fully) capture as profits (Spence (1976), Mankiw and Whinston (1986)). The supply diversity and business-stealing effects thus work in opposite directions. Whether the business-stealing or diversity effect dominates depends on the shape of the demand curve and the nature of post-entry competition.

We disentangle the two effects via simulations of two counterfactuals. In the Planner's benchmark, we solve for the optimal dynamic entry path chosen by a social planner who maximizes total surplus. In the Coordination benchmark, we search for the optimal entry path chosen by a planner who maximizes producer surplus but takes post-entry competition as given.[45] In the first scenario, both business-stealing and diversity effects are internalized, whereas in the second scenario, only the business-stealing effect is accounted for by the

[45] In the coordination benchmark, the planner does not coordinate output across active firms but only coordinate their entry decisions taking the oligopolistic post-entry competition as given.

planner. We note that the solutions to these dynamic optimization problems are greatly simplified by our assumption that only one potential entrant has an opportunity to enter every quarter. This assumption reduces the dimensionality of the state space for the social planner to be the same as that of a player in the dynamic game.[46]

The results of this counterfactual exercise are shown in Table 9. Relative to the market outcome, total surplus under the Planner's benchmark is on average 6% higher. In the coordination benchmark, the expected number of cables in 2021 is 16% lower compared to the market outcome, producer surplus increases by 16% as business-stealing is internalized. However, consumer surplus decreases by 8%. We compare the size of the welfare gaps due to business-stealing and diversity effects: distortions due to diversity effects (in absolute value) amount to between 83% and 101% of the distortions due to business-stealing. For most markets, we find that business-stealing dominates leading to moderately excessive entry rates.

8 Conclusion

Supply disruptions and the ensuing diversification sought by buyers are important features in many industries. This paper empirically assesses their quantitative impact on an industry's dynamic evolution in the context of the global internet backbone. To do so, we build a dynamic oligopoly game of entry by cable operators who supply bandwidth to buyers with preferences for network diversity. The model is estimated using a novel data on used bandwidth, bandwidth prices, and cable characteristics.

We find that supplier diversification accounts for a large share of entry rates and surplus created over the sample period 2005-2021. Moreover, we quantify entry bias due to entrants' inability to capture the benefits of diversity. Indeed, when buyers have a preference for network diversity, a marginal entrant, by increasing diversity, increases surplus but does not fully capture this gain in profits. We find that distortions due to this diversity effects are of similar magnitude as distortions due to standard business-stealing effect.

The proposed empirical strategy may be useful in other contexts characterized by bottlenecks (e.g., energy transportation, semiconductor manufacturing) or when evaluating the impact of mergers in this type of industries. In this paper it is applied to the global internet backbone, an industry characterized by frequent and costly supply disruptions and buyer preference for diverse physical paths. Studying this industry is important in its own right,

[46]With simultaneous moves of multiple players every period, the social planner would need to choose actions based on the entire vector of players' private information shocks. Under the one-mover assumption, the social planner's decision only depends on a single potential entrant's private shock.

Table 9: Counterfactual outcomes under the planner and coordination benchmarks

	DGP		CF		CF/DGP	
	Mean	Median	Mean	Median	Mean	Median
Social Planner benchmark						
Entry rates						
Number of cables in 2005-Q1 (data)	1.56	1.00				
Number of cables in 2021-Q4 (data)	3.20	3.00				
Expected number of cables in 2021-Q4	3.17	2.50	3.14	2.44	0.96	0.95
Expected probability of entry	0.033	0.014	0.022	0.000	0.644	0.012
Welfare (2005-Q1 to 2021-Q4) in millions of US$						
Consumer Surplus	508.87	59.77	537.71	56.74	1.00	0.98
Producer Surplus	385.63	279.82	405.25	270.52	1.07	1.06
Total Surplus	894.51	385.23	942.96	416.78	1.06	1.05
Consumer Surplus / Total Surplus	0.30	0.22	0.29	0.19	0.94	0.95
Coordination benchmark						
Entry rates						
Number of cables in 2005-Q1 (data)	1.56	1.00				
Number of cables in 2021-Q4 (data)	3.20	3.00				
Expected number of cables in 2021-Q4	3.17	2.50	2.72	2.00	0.84	0.87
Expected probability of entry	0.033	0.014	0.003	0.000	0.180	0.000
Welfare (2005-Q1 to 2021-Q4) in millions of US$						
Consumer Surplus	508.87	59.77	421.20	56.57	0.92	0.95
Producer Surplus	385.63	279.82	441.27	333.25	1.16	1.09
Total Surplus	894.51	385.23	862.46	416.78	1.04	1.02
Consumer Surplus / Total Surplus	0.30	0.22	0.26	0.19	0.89	0.92

Note: The first two columns correspond to the data generating process (DGP) and show outcomes under the equilibrium played in the data. The following two columns show counterfactual (CF) outcomes under the social planner and coordination benchmarks. The last two columns show the ratio of the counterfactual outcomes to DGP outcomes. The variable "expected probability of entry" is the average over all periods of the expected probability of entry given equilibrium state transitions and choice probabilities.

as it provides insights into the infrastructure and physical connections that support the internet, a central underpinning of modern communication and commerce.

References

Aguirregabiria, Victor, and Mathieu Marcoux. 2021. "Imposing equilibrium restrictions in the estimation of dynamic discrete games." *Quantitative Economics* 12 1223–1271.

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica* 75 (1): 1–53.

Aguirregabiria, Victor, and Junichi Suzuki. 2014. "Identification and counterfactuals in dynamic models of market entry and exit." *Quantitative Marketing and Economics* 12 (3): 267–304.

Babich, Volodymyr, Apostolos N. Burnetas, and Peter H. Ritchken. 2007. "Competition and diversification effects in supply chains with supplier default risk." *Manufacturing and Service Operations Management* 9 123–146.

Bajari, Patrick, Victor Chernozhukov, Han Hong, and Denis Nekipelov. 2015. "Identification and Efficient Semiparametric Estimation of a Dynamic Discrete Game."

Baldwin, Richard, and Rebecca Freeman. 2022. "Risks and Global Supply Chains: What We Know and What We Need to Know." *Annual Review of Economics* 14 153–180.

Barrot, Jean Noël, and Julien Sauvagnat. 2016. "Input specificity and the propagation of idiosyncratic shocks in production networks." *Quarterly Journal of Economics* 131 1543–1592.

Berry, Steven, Alon Eizenberg, and Joel Waldfogel. 2016. "Optimal product variety in radio markets." *The RAND Journal of Economics* 47 (3): 463–497.

Besen, Stanley, Paul Milgrom, Bridger Mitchell, and Padmanabhan Srinagesh. 2001. "Advances in Routing Technologies and Internet Peering Agreements." *American Economic Review* 91 (2): 292–296.

Blum, Bernardo S., and Avi Goldfarb. 2006. "Does the internet defy the law of gravity?" *Journal of International Economics* 70 384–405.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2019. "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tohoku Earthquake." *Review of Economics and Statistics* 101 60–75.

Caillaud, Bernard, and Bruno Jullien. 2003. "Chicken Egg: Competition among Intermediation Service Providers." *The RAND Journal of Economics* 34 309–328.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. 2021. "Supply chain disruptions: Evidence from the great east japan earthquake." *Quarterly Journal of Economics* 136 1255–1321.

Castro-Vincenzi, Juanma. 2022. "Climate Hazards and Resilience in the Global Car Industry."

Collard-Wexler, Allan. 2013. "Demand Fluctuations in the Ready-Mix Concrete Industry." *Econometrica* 81 1003–1037.

Conte, Maddalena, Pierre Cotterlaz, and Thierry Mayer. 2021. "The CEPII gravity database." *CEPII: Paris, France.*

Crémer, Jacques, Patrick Rey, and Jean Tirole. 2000. "Connectivity in the Commercial Internet." *The Journal of Industrial Economics* 48 433–472.

Dixit, Robert K, Avinash K Dixit, and Robert S Pindyck. 1994. *Investment under uncertainty.* Princeton university press.

Dormon, Bob. 2016. "How the Internet works." May, arstechnica.com [Online; posted 26-May-2016].

Draganska, Michaela, Michael Mazzeo, and Katja Seim. 2009. "Beyond plain vanilla: Modeling joint product assortment and pricing decisions." *Quantitative Marketing and Economics* 7 105–146.

Eizenberg, Alon. 2014. "Upstream innovation and product variety in the U.S. home PC market." *Review of Economic Studies* 81 1003–1045.

Elliott, Matthew, Benjamin Golub, and Matthew V. Leduc. 2022. "Supply Network Formation and Fragility." *American Economic Review* 112 2701–2747.

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

Federal Communications Commission. 2015. "FCC Record 15-119: A Comprehensive Compilation of Decisions, Reports, Public Notices, and Other Documents of the Federal Communications Commission of the United States."

Garbade, Kenneth D., and William L. Silber. 1978. "TECHNOLOGY, COMMUNICATION AND THE PERFORMANCE OF FINANCIAL MARKETS: 1840-1975." *The Journal of Finance* 33 819–832.

Greenstein, Shane. 2015. *How the internet became commercial: Innovation, privatization, and the birth of a new network.* Princeton University Press.

Greenstein, Shane, and Tommy Pan Fang. 2020. "Where the Cloud Rests: The Economy Geography of Data Centers.." *Harvard Business School Working Paper, No. 21-042.*

Grossman, Gene M., Elhanan Helpman, and Hugo Lhuillier. 2023. "Supply Chain Resilience: Should Policy Promote International Diversification or Reshoring?" *Journal of Political Economy.*

Hjort, Jonas, and Jonas Poulsen. 2019. "The arrival of fast internet and employment in Africa †." *American Economic Review* 109 1032–1079.

Hoag, Christopher. 2006. "The Atlantic Telegraph Cable and Capital Market Information Flows." *The Journal of Economic History* 66 342–353.

Igami, Mitsuru. 2017. "Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998." *Journal of Political Economy* 125 (3): 798–847.

Igami, Mitsuru, and Kosuke Uetake. 2020. "Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996-2016." *Review of Economic Studies* 87 2672–2702.

ISCPC. 2022. "Cableships of the World." February, iscpc.org [Online; udpated 11-February-2022].

Jeon, Jihye. 2022. "Learning and investment under demand uncertainty in container shipping." *The RAND Journal of Economics* 53 (1): 226–259.

Jiang, Bomin, Daniel Rigobon, and Roberto Rigobon. 2022. "From Just-in-Time, to Just-in-Case, to Just-in-Worst-Case: Simple Models of a Global Supply Chain under Uncertain Aggregate Shocks." *IMF Economic Review* 70 141–184.

Judd, Kenneth L. 1998. *Numerical Methods in Economics.* Volume 1. of MIT Press Books, The MIT Press, , https://ideas.repec.org/b/mtp/titles/0262100711.html.

Kalouptsidi, Myrto. 2014. "Time to Build and Fluctuations in Bulk Shipping." *American Economic Review* 104 564–608.

Kalouptsidi, Myrto, Yuichi Kitamura, Eduardo Souza-Rodrigues, and Lucas Lima. 2021a. "Counterfactual Analysis for Structural Dynamic Discrete Choice Models." (June): .

Kalouptsidi, Myrto, Paul Scott, and Eduardo Souza-Rodrigues. 2021b. "Identification of Counterfactuals in Dynamic Discrete Choice Models." *Quantitative Economics.*

Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole. 2001. “Internet Peering.” *The American Economic Review* 91 287–291.

Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole. 2003. “Internet Interconnection and the Off-Net-Cost Pricing Principle.” *The RAND Journal of Economics* 34 370–390.

Loertscher, Simon, and Michael H Riordan. 2019. “Make and buy: Outsourcing, vertical integration, and cost reduction.” *American Economic Journal: Microeconomics* 11 (1): 105–123.

MacKie-Mason, Jeffrey K., and Hal Varian. 1994. “Economic FAQs About the Internet.” *Journal of Economic Perspectives* 8 (3): 75–96.

Magnac, Thierry, and David Thesmar. 2002. “Identifying Dynamic Discrete Decision Processes.” *Econometrica* 70 (2): 801–816.

Malphrus, Steve. 2009. “Keynote Address, Reliability of Global Undersea Communication Cables Infrastructure (ROGUCCI), Summit Dubai, UAE.”

Mankiw, N Gregory, and Michael D Whinston. 1986. “Free Entry and Social Inefficiency.” *The RAND Journal of Economics* 17 48–58.

Pakes, Ariel. 2017. “Empirical Tools and Competition Analysis: Past Progress and Current Problems.” *International Journal of Industrial Organization* 241–266.

Pesendorfer, Martin, and Philipp Schmidt-Dengler. 2008. “Asymptotic least squares estimators for dynamic games.” *Review of Economic Studies* 75 (3): 901–928.

Pesendorfer, Martin, and Philipp Schmidt-Dengler. 2010. “Sequential Estimation of Dynamic Discrete Games: A Comment.” *Econometrica* 78 (2): 833–842.

Reverdy, Dominique, and Ivan Skenderoski. 2015. “Submarine Cables: Structuring and Financing Options.” *Salience Whitepaper.*

Rust, John. 1994. “Structural Estimation Decision Processes.” *Handbook of Econometrics* Chapter 51 3139–3139.

Rust, John. 2000. “Parametric Policy Iteration : An Efficient Algorithm for Solving Multidimensional DP Problems?.”Technical report.

Seixas, Pedro. 2015. “Broadband costing and pricing - Approaches and best practices.” *International Telecommunication Union.*

Spence, Michael. 1976. "Product Selection, Fixed Costs, and Monopolistic Competition."

Steinwender, Claudia. 2018. "Real Effects of Information Frictions: When the States and the Kingdom Became United." *American Economic Review* 108 657–696.

Sweeting, Andrew. 2013. "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry." *Econometrica* 81 (5): 1763–1803.

Swinhoe, Dan. 2021. "What is a Subsea Cable?." August, datacenterdynamics.com [Online; posted 26-August-2021].

Tang, Sammi Yu, and Panos Kouvelis. 2011. "Supplier diversification strategies in the presence of yield uncertainty and buyer competition." *Manufacturing and Service Operations Management* 13 439–451.

Tauchen, George. 1986. "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions." *Economics Letters* 20 177–181.

TeleGeography. 2022. "Product Documentation." teleogeography.com .

The Economist. 2002. "The great telecoms crash." July, economist.com [Online; posted 18-July-2002].

The Guardian. 2016. "Buried at sea: the companies cashing in on abandoned cables." December, theguardian.com [Online; posted 14-December-2016].

Villas-Boas, Sofia Berto. 2007. "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data." *The Review of Economic Studies* 74 (2): 625–652, http://www.jstor.org/stable/4626153.

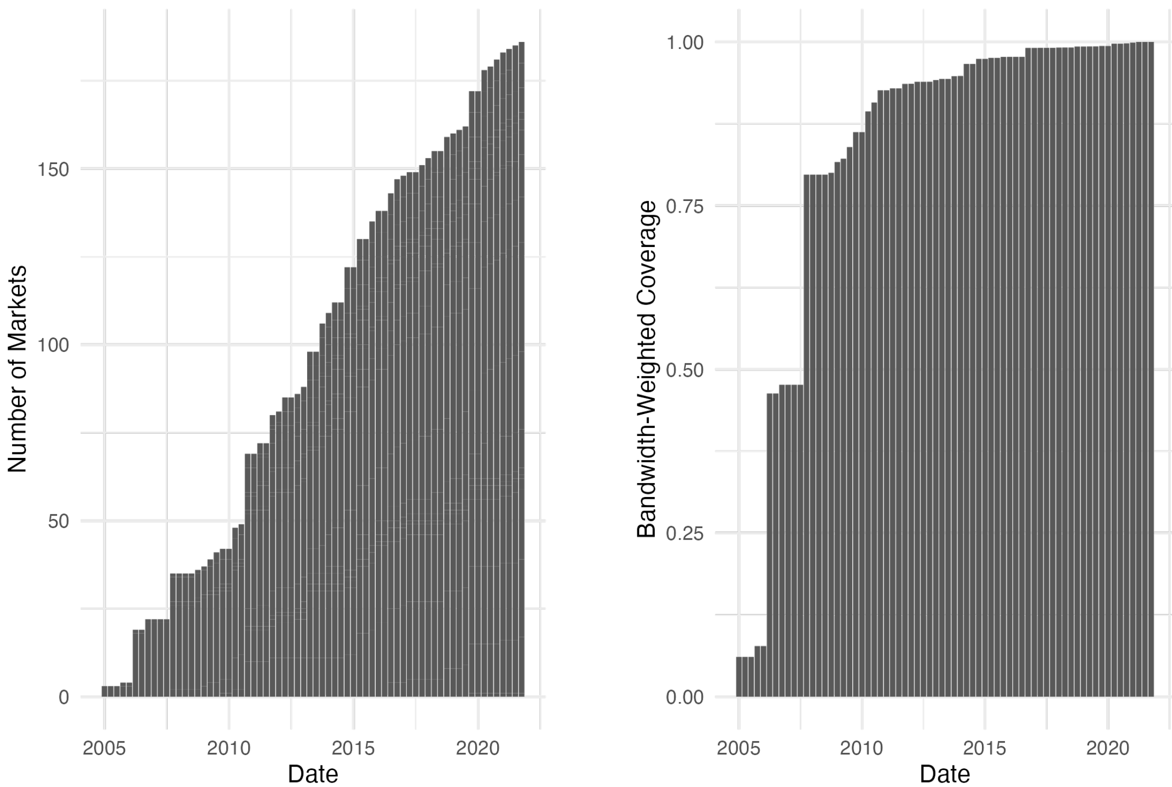
Williams, Kevin R. 2022. "The Welfare Effects of Dynamic Pricing: Evidence From Airline Markets." *Econometrica* 90 831–858.

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

A Additional Information on the Market Definition

This section provides details on the markets contained in the TeleGeography datasets that are used in estimation. As discussed in Section 4.3, we define a market as a country pair. Our estimation panel is not balanced; new cables are continually built, connecting new country pairs resulting in new markets appearing in the data. This is illustrated in the left panel of Figure A1; a detailed breakdown by region-pair is shown in Figure A2.

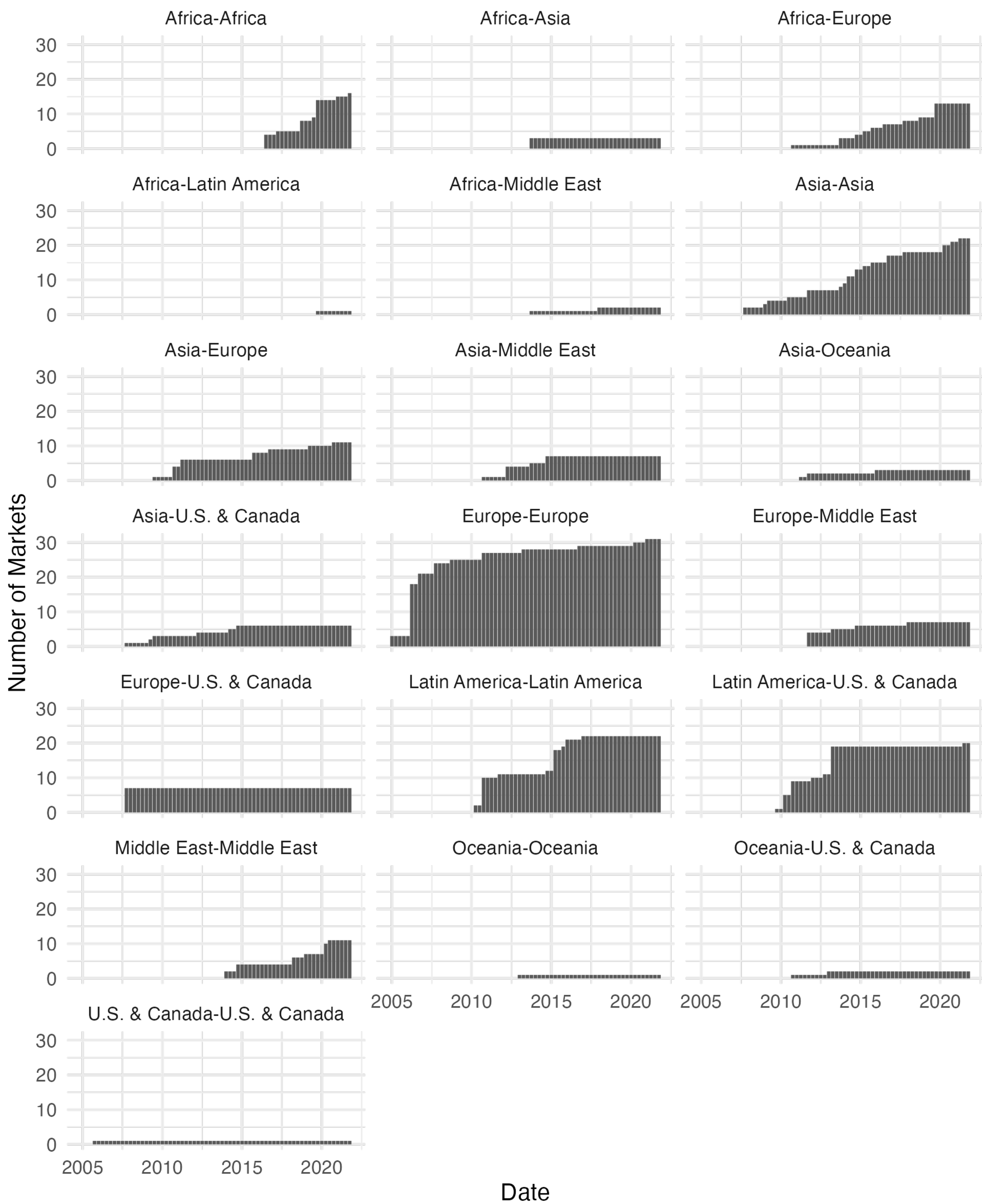
Figure A1: Cumulative Number of Markets (left) and Cumulative Bandwidth Coverage (right) in the Estimation Panel



Notes: The left figure shows the cumulative number of distinct country-to-country markets for which both price and used bandwidth data are available. The right figure shows final-period bandwidth-weighted cumulative coverage in the estimation panel.

However, on a bandwidth-weighted basis, the estimation panel is well balanced, as is shown in Figure A1 (right panel) and A3. The right panel of Figure A1 plots the final-period bandwidth-weighted cumulative coverage of the estimation panel (final-period bandwidth is used to account for overall bandwidth growth over the data). It can be seen that beginning in the fourth quarter of 2007 the panel's coverage reached 80% of aggregate final-period bandwidth and thereafter remained substantially complete. Figure A3 shows similar plots broken down by region pairs: the panel's coverage is well balanced for those markets with

Figure A2: Cumulative Number of Markets in the Estimation Panel by Region-to-Region Pair



Notes: This figure shows the cumulative number of distinct country-to-country markets, for which both price and used bandwidth data are available, broken down by region pair.

one landing point in Europe, North America or Latin America. Growth in African, Asian, and Middle Eastern markets drive most of the estimation panel's imbalance. The difference between these two Figures and Figures A1 (left panel) – A2 illustrates the out-sized role that markets with one landing point in Europe or North America play in the global statistics.

We provide further details on the markets included in our sample in Tables A1 and A2. Table A1 contains information on the largest 15 markets (by bandwidth) in the estimation panel, showing the first and last period for which information (on prices and bandwidth used) is available in the data.

Table A1: Top 15 Markets

Market	First Period	Final Period	Observations	Bandwidth
Germany-United Kingdom	2006-Q2	2021-Q4	63	133,950.7
France-Germany	2006-Q2	2021-Q4	63	131,869.1
Germany-Netherlands	2006-Q2	2021-Q4	63	114,501.2
United Kingdom-United States	2007-Q4	2021-Q4	57	110,773.3
France-United Kingdom	2006-Q2	2021-Q4	63	103,257.7
Netherlands-United Kingdom	2006-Q2	2021-Q4	63	98,091.4
France-Spain	2006-Q2	2021-Q4	63	96,661.1
Japan-United States	2007-Q4	2021-Q4	57	93,189.8
Denmark-United States	2007-Q4	2021-Q4	12	91,767.4
Ireland-United Kingdom	2007-Q4	2021-Q4	57	85,333.5
Spain-United States	2007-Q4	2021-Q4	20	71,005.0
Denmark-Sweden	2005-Q1	2021-Q4	68	61,609.9
Denmark-Germany	2005-Q1	2021-Q4	37	59,898.3
France-United States	2007-Q4	2021-Q4	57	58,693.1
Ireland-United States	2007-Q4	2021-Q4	55	53,448.6

Note: This table presents some details on the top 15 markets in the data, in terms of used bandwidth in the final period. Columns 2 and 3 contain the first and last period each market appears in our estimation panel. Column 4 contains the number of periods for each market in the final estimation panel. Column 5 contains the used bandwidth in the final period.

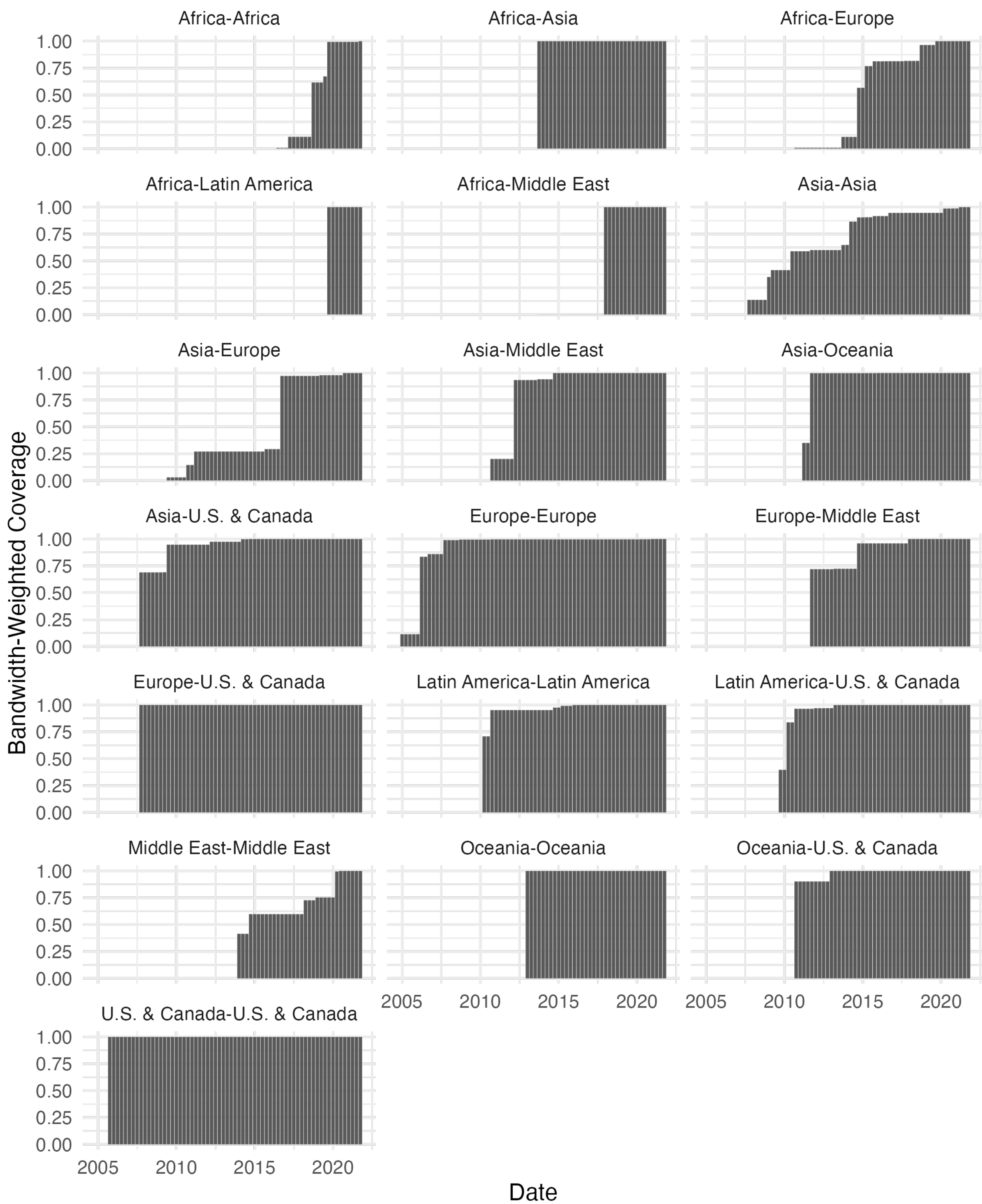
Table A2 shows similar data broken down by region pair. For reasons of brevity, only the largest 3 markets in each region pair (by bandwidth) are shown.

B Extensions

B.1 Time-to-Build

In this section, we extend the industry model to account for the fact that cables generally take longer than a single period (3 calendar months) to build. A full specification of this

Figure A3: Cumulative Bandwidth Coverage in the Estimation Panel by Region-to-Region Pair



Notes: This figure shows final-period bandwidth-weighted cumulative coverage in the estimation panel, broken down by region-to-region pair.

Table A2: Top 3 Markets by Region Pair

Market	First Period	Final Period	Log Bandwidth	Market	First Period	Final Period	Log Bandwidth
Africa-Africa				**Europe-Europe**			
Kenya-South Africa	2018-Q4	2021-Q4	827.4	Germany-U.K.	2006-Q2	2021-Q4	133,950.7
Ghana-Nigeria	2017-Q2	2021-Q4	223.6	France-Germany	2006-Q2	2021-Q4	131,869.1
Angola-South Africa	2019-Q4	2021-Q3	204.4	Germany-Netherlands	2006-Q2	2021-Q4	114,501.2
Africa-Asia				**Europe-Middle East**			
Egypt-India	2013-Q4	2015-Q1	23.1	France-U.A.E.	2011-Q4	2021-Q4	3,516.6
India-Kenya	2013-Q4	2019-Q4	11.0	France-Saudi Arabia	2011-Q4	2021-Q4	3,366.5
India-Tanzania	2013-Q4	2015-Q4	0.7	France-Oman	2014-Q4	2021-Q4	2,673.1
Africa-Europe				**Europe-U.S. & Canada**			
Egypt-France	2014-Q4	2021-Q4	5,885.6	U.K.-U.S.	2007-Q4	2021-Q4	110,773.3
Nigeria-U.K.	2015-Q2	2021-Q4	2,596.2	Denmark-U.S.	2007-Q4	2021-Q4	91,767.4
France-South Africa	2018-Q4	2021-Q1	1,911.4	Spain-U.S.	2007-Q4	2021-Q4	71,005.0
Africa-Latin America				**Latin America-Latin America**			
Angola-Brazil	2019-Q4	2021-Q4	164.0	Argentina-Chile	2010-Q2	2021-Q4	6,546.6
				Argentina-Brazil	2010-Q2	2021-Q4	5,957.5
				Chile-Peru	2010-Q4	2021-Q4	2,211.2
Africa-Middle East				**Latin America-U.S. & Canada**			
Kenya-U.A.E.	2018-Q1	2020-Q3	282.5	Brazil-U.S.	2009-Q4	2021-Q4	45,624.8
Egypt-Oman	2013-Q4	2015-Q1	0.2	Mexico-U.S.	2010-Q2	2021-Q4	29,696.1
				Chile-U.S.	2010-Q2	2021-Q4	11,423.4
Asia-Asia				**Middle East-Middle East**			
Japan-Singapore	2009-Q1	2021-Q4	32,841.5	Bahrain-U.A.E.	2014-Q1	2021-Q2	361.5
India-Singapore	2010-Q3	2021-Q4	27,004.4	Kuwait-U.A.E.	2014-Q4	2021-Q2	178.3
Indonesia-Singapore	2014-Q2	2021-Q4	24,756.7	Qatar-U.A.E.	2014-Q1	2021-Q2	154.5
Asia-Europe				**Oceania-Oceania**			
France-India	2016-Q4	2021-Q4	22,429.3	Australia-New Zealand	2013-Q1	2021-Q4	3,827.7
France-Singapore	2011-Q2	2021-Q4	4,011.1				
India-U.K.	2010-Q4	2021-Q4	3,674.8				
Asia-Middle East				**Oceania-U.S. & Canada**			
Saudi Arabia-Singapore	2012-Q2	2021-Q4	1,074.8	Australia-U.S.	2010-Q4	2021-Q4	12,133.2
Singapore-U.A.E.	2012-Q2	2021-Q4	999.8	New Zealand-U.S.	2013-Q1	2021-Q4	1,313.5
India-U.A.E.	2010-Q4	2021-Q4	567.8				
Asia-Oceania				**U.S. & Canada-U.S. & Canada**			
Australia-Singapore	2011-Q4	2021-Q4	3590.6	Canada-U.S.	2005-Q4	2021-Q4	32,980.7
Australia-Japan	2011-Q2	2021-Q4	1933.4				
Australia-Indonesia	2016-Q1	2018-Q2	5.6				
Asia-U.S. & Canada							
Japan-U.S.	2007-Q4	2021-Q4	93,189.8				
Singapore-U.S.	2009-Q3	2021-Q4	34,663.6				
South Korea-U.S.	2012-Q2	2021-Q4	3,774.9				

Note: This table presents some details on the top 3 markets by region pair, in terms of used bandwidth in the final period. Columns 2 and 3 contain the first and last period each market appears in our estimation panel. Column 4 contains the (log) used bandwidth in the final period.

extension would include two separate features: First, a function mapping cable and market characteristics (e.g. geography and cable length) to a time-to-build; second, a modification of the state space to account for the number of cables currently being built and how many periods remain (in expectation) to their completion. While this second feature is straightforward in principle, in practice, it greatly increases the size of the state space making the empirical implementation challenging. To illustrate the problem, with a time-to-build of two years, the entry probability for a cable in a given quarter would depend on how many cables had begun being built in each of the prior seven calendar quarters (representing seven additional dimensions). This full specification is therefore beyond the scope of the present paper.

Instead, we use a characterization of time-to-build that captures the underlying economics without imposing an insurmountable burden on estimation. We assume that each new cable entrant is added to the pool of cables under construction; the cable then emerges from the construction state to the market competition state at a constant exogenous market-specific rate ζ_m. Thus, when forming expectations, a potential entrant need only keep track of the number of cables currently under construction, denoted e_{mt}. The expected number of new competitors finishing construction in the next period is given by $\zeta_m e_{mt}$ (expectation of a binomial distribution $\mathcal{B}(e_{mt}, \zeta_m)$), and if the cable operator decides to begin construction, it will expect to wait $1/\zeta_m$ periods before completing construction and starting to operate.

In this setting, the only part of the industry model that needs to be modified are the dynamics. The bandwidth demand model, the static competition game, and per-period profits are unchanged, as are the transitions for demand and cost states d_{mt} and h_{mt}.

First, the industry state becomes

$$\mathcal{M}_{mt} = (n_{mt}, e_{mt}, d_{mt}, h_{mt})\,, \tag{31}$$

where the additional component e_{mt} represents the number of under-construction cables in market m in period t. We assume that under-construction cables do not make any choices, but simply wait for their construction process to finish, with constant exogenous probability ζ_m in each period (set such that the expected wait time under construction is 2 years).

Second, while the profit Equation (5) is unchanged, we note that entry costs EC_{mt} are paid in the period that the cable enters construction, and variable profits $VP(\mathcal{M}_t)$ are zero until construction is completed.

Third, the choice-specific value functions for the potential entrant then become

$$v^{\alpha}_{i,t}(a_{it}, \mathcal{M}_t) = \begin{cases} -EC_{mt} + \beta\,\mathbb{E}_t\left[V^{\alpha,e}_{i,t+1}\left(\mathcal{M}_{t+1}\right)\right] & \text{if} \quad a_{it} = 1 \\ 0 & \text{if} \quad a_{it} = 0 \end{cases} \tag{32}$$

where the (next-period) value function for an under-construction cable $V^{\alpha,e}_{i,t+1}$ is

$$V^{\alpha,e}_{i,t+1}\left(\mathcal{M}_{t+1}\right) = \zeta_m \beta\, \mathbb{E}_{t+1}\left[V^{\alpha}_{i,t+2}\left(\mathcal{M}_{t+2}\right)\right] + (1-\zeta_m)\beta\, \mathbb{E}_{t+1}\left[V^{\alpha,e}_{i,t+2}\left(\mathcal{M}_{t+2}\right)\right] \tag{33}$$

and $V^{\alpha}_{i,t+2}$ is the value function of an incumbent cable operator. The estimates of δ_ϵ remain qualitatively similar to the baseline, although their magnitude is smaller reflecting the fact that the continuation value from entering (Equation (32)) is lower with time-to-build.

B.2 Exogenous cable retirement

As in the previous section, we can introduce an alternative extension to the dynamic game: the retirement of cables after a finite time horizon. This addition to the model brings it closer to what is likely to be the long-run equilibrium in the industry (though as discussed in Section 4, not the years covered in our data). Industry consensus is that cables have roughly a 25 year lifespan (this is the length of a typical FCC license for a cable landing in the US), though some cables have been in service for longer and have received regulatory extensions.

As in Section B.1, a fully-specified version of cable retirement would lead to a curse of dimensionality. For example, under the simple assumption that each cable would retire after 25 years, a cable considering entry in a market with three incumbent cables would have a different values from entering depending on the age distribution of the three incumbents. The value of entering a market with a 1-year old incumbent, a 5-year old incumbent, and a 7-year old incumbent differs from the value of entering the same market with a 22 year-old incumbent, a 23-year old incumbent and a 24-year old incumbent cables. The value function would need to depend not only on the number of incumbent cables but their age distribution as well.

We proceed with a simplified version of cable retirement that keeps estimation tractable. We assume that each cable faces an exogenous (and known) probability of retirement in each period, which we denote χ.[47] This corresponds to each cable having an expected lifespan of $1/\chi$ periods. Cable retirement is a terminal action. A retired cable is replaced by a new potential entrant.

Under this assumption, the model described in Section 5 can be used with a few minor modifications. The dynamic state remains

$$\mathcal{M}_{mt} = \left(n_{mt}, d_{mt}, h_{mt}\right), \tag{34}$$

[47] In principle, the exit probability can be market-specific (χ_m), but with only a handful of exit events in our data, we cannot empirically tie an exit probability to other market characteristics.

and the introduction of the exogenous exit rate χ alters the expectation of future competition, as (without entry) $\mathbb{E}_t[n_{t+1}] = (1-\chi)n_t$.[48] The choice-specific value function for incumbent firms becomes:

$$v_{i,t}^{\alpha}(a_{it}, \mathcal{M}_t) = \begin{cases} \pi_i(a_{it}, \mathcal{M}_t) + \beta(1-\chi)\,\mathbb{E}_t\left[V_{i,t+1}^{\alpha}\left(\mathcal{M}_{t+1}, \boldsymbol{\epsilon}_{i,t+1}\right)\right] + \beta\chi EV_{mt} & \text{if} \quad a_{it} = 1 \\ \pi_i(a_{it}, \mathcal{M}_t) + EV_{mt} & \text{if} \quad a_{it} = 0 \end{cases} \tag{35}$$

where it can be seen that even if an incumbent chooses to remain active ($a_{it} = 1$), they will exogenously exit and receive EV_{mt} with probability χ. We set χ so that the expected lifespan of a cable is 25 years. This specification does not significantly change the estimates of δ_ϵ relative to the baseline (e.g., the 1-PML estimate $\widehat{\delta_\epsilon}$ equals 0.234 under $\delta_{FC} = 2\%$).

B.3 Asymmetries in costs

In this section, we allow for heterogeneity in marginal costs across cables with differing vintages (i.e. built at different times). We consider a setup with J different vintages. Suppressing the market and time subscripts for legibility, the number of incumbent cables belonging to vintage j is denoted n_j, with $n = \sum_J n_j$. The marginal costs of bandwidth for each type are likewise denoted c_j; we consider a type-symmetric equilibrium of the static Cournot game and denote the (per-firm) quantity produced by firms of type j by q_j, with the market-period identity $Q = \sum_j n_j q_j$.

A firm of type j chooses the optimal amount of bandwidth to supply by satisfying the first-order condition (omitting the market and time subscripts)

$$p - c_j + q_j \frac{\partial p}{\partial q_j} = 0\,,$$

which can be rewritten as

$$\frac{p - c_j}{p} = -\frac{q_j}{Q}\frac{Q}{p}\frac{\partial p}{\partial q_j} = -\frac{q_j}{Q}\frac{1}{\epsilon}\,,$$

where ε is the price elasticity of demand. Summing over all firms yields

$$\sum_{j=1}^{J} n_j \left(\frac{p - c_j}{p}\right) = -\frac{1}{\epsilon}\,, \tag{36}$$

which can be rearranged to express equilibrium prices as a function of the number of firms

[48] The state space is larger under this extension because n_t can potentially decrease, whereas under the baseline model without exit, n_t is weakly increasing over time.

and marginal costs by type:

$$p = \frac{1}{n + \varepsilon^{-1}} \sum_{j=1}^{J} n_j c_j \, . \tag{37}$$

Noting that p, n_j, and ε are either observed in the data or estimated objects, one can then recover the marginal costs by firm type $\{c_j\}_{j \in J}$ by exploiting the cross-sectional and time-series variation in the vintage distribution of incumbent cables.

In practice, we set J to 2 and assign cables to vintages with the following cutoff

$$j(i) = \begin{cases} 1 & \text{if} \quad t_i \leq \widetilde{t} \\ 2 & \text{if} \quad t_i > \widetilde{t} \end{cases}$$

where t_i is the period cable i entered service, and $\widetilde{t}$ is a cut-off period. We assume that marginal costs are the sum of market-specific δ_m, period-specific δ_t, and vintage-specific fixed effects $\widetilde{\delta}_j$, for $j \in \{1, 2\}$. This allows markets with different physical characteristics (e.g. distance, propensity for cable faults) to have different marginal costs while restricting the within-market cost difference between different vintages of cables to be common across all markets. Kalouptsidi (2014) uses a similar approach for ship ages. We estimate the following equation via ordinary least squares (denoting the error term by ν_{mt})

$$p_{mt} \left(\frac{n_{mt} + \epsilon^{-1}}{n_{mt}} \right) = \delta_m + \delta_t + \frac{n_{1,mt}}{n_{mt}} \widetilde{\delta}_1 + \frac{n_{2,mt}}{n_{mt}} \widetilde{\delta}_2 + \nu_{mt} \tag{38}$$

Table A3 shows the estimation results for different choices of the cutoff quarter $\widetilde{t}$ (e.g., 2010 corresponds to 2009-Q4). We do not find evidence of significant heterogeneity in costs across cables of different vintages, once we control for market and period fixed effects. This is related to the discussion in Section 6.4 (see footnote 32) suggesting that many technological upgrades (that affect the marginal cost of lighting additional bandwidth) can be installed on equipment at the landing stations and, therefore, benefit incumbent cables as well as new entrants.

C Supplementary Tables and Figures

In Figure A4, we use simulations to illustrate the comparative static properties of the inverse demand curves (derived from Equation (4)) that arise from the model outlined in Section 2. The no-disruption case is also plotted. Except where otherwise indicated, the parameters used are: $a = 10$, $b = 0.125$, $\mu = 1$, $\sigma^2 = 0.5$, $n = 1$. The use of $\mu = 1$ is to ensure comparability to the no-disruption case; effects illustrated would be increased with $\mu < 1$.

In the first sub-figure, the effect of n, or the number of firms in the market, on the inverse

Table A3: Heterogeneity in marginal costs

	Dependent variable: $p_{mt}\left(\frac{n_{mt}+\epsilon^{-1}}{n_{mt}}\right)$		
	$\widetilde{t}=2005$	$\widetilde{t}=2010$	$\widetilde{t}=2015$
$n_{1,mt}/n_{mt}$	13378.0	13443.2	12761.0
	(2403.8)	(2324.4)	(2235.4)
$n_{2,mt}/n_{mt}$	12849.2	12600.8	14379.9
	(3495.7)	(3559.8)	(3879.8)
Market FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
R^2	0.776	0.776	0.776
Adjusted R^2	0.763	0.763	0.763
Within R^2	0.009	0.009	0.009
Observations	4,777	4,777	4,777

Note: The unit of observation is the market (country pair) by quarter. Standard errors (in parenthesis) are clustered at the market level. The depend variable is expressed in $US.

demand curve is represented. The figure plots inverse demand curves for $n = 1, 2, 4, 8$, holding other parameters fixed. It also plots the no-disruption case as a benchmark.

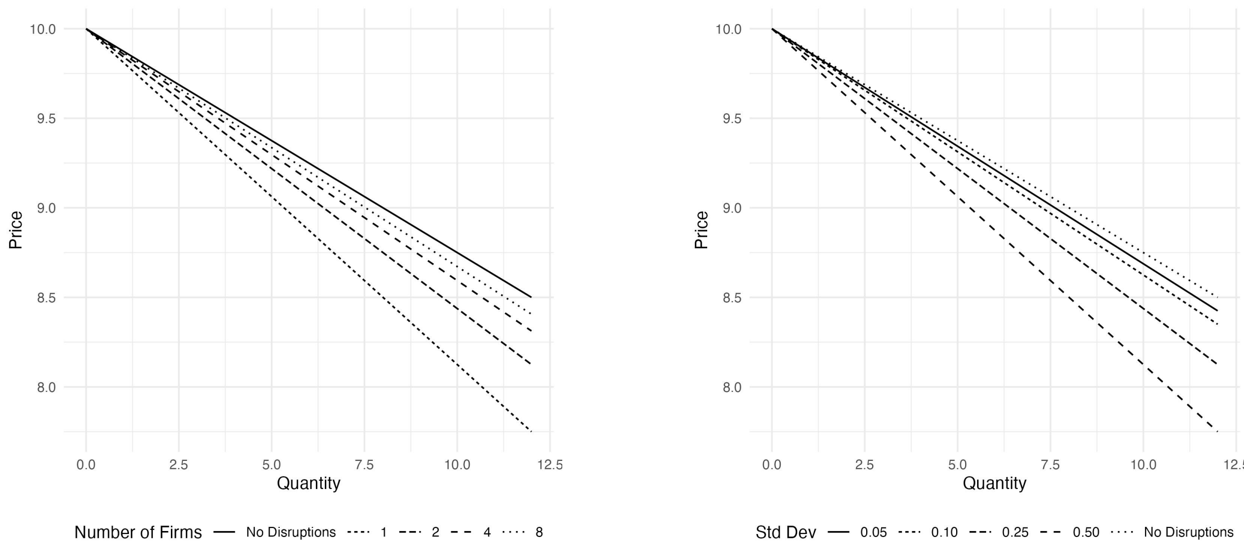
The second sub-figure illustrates the *ceteris paribus* effect of changes in σ^2 on $p(q)$. The figure plots inverse demand curves for $\sigma^2 = 0.05, 0.10, 0.25, 0.5$, holding other parameters fixed. It also plots the no-disruption case as a benchmark.

Together, these two sub-figures demonstrate the first-order importance of the parameters n and σ^2 on mediating $p(q)$ in a market with disruption risk.

Table A4: Statistics on Cable Ownership

	Percentile							
	Minimum	5th	25th	50th	75th	95th	Maximum	Mean
Owners per Cable	1	1	1	1	2	9.00	51	2.5
Cables per Owner	1	1	1	1	3	11.55	26	2.9
Cables per Owner-Market	1	1	1	1	1	2.00	7	1.1

Figure A4: Simulated Comparative Statics



(a) Comparative Statics in n (b) Comparative Statics in σ^2

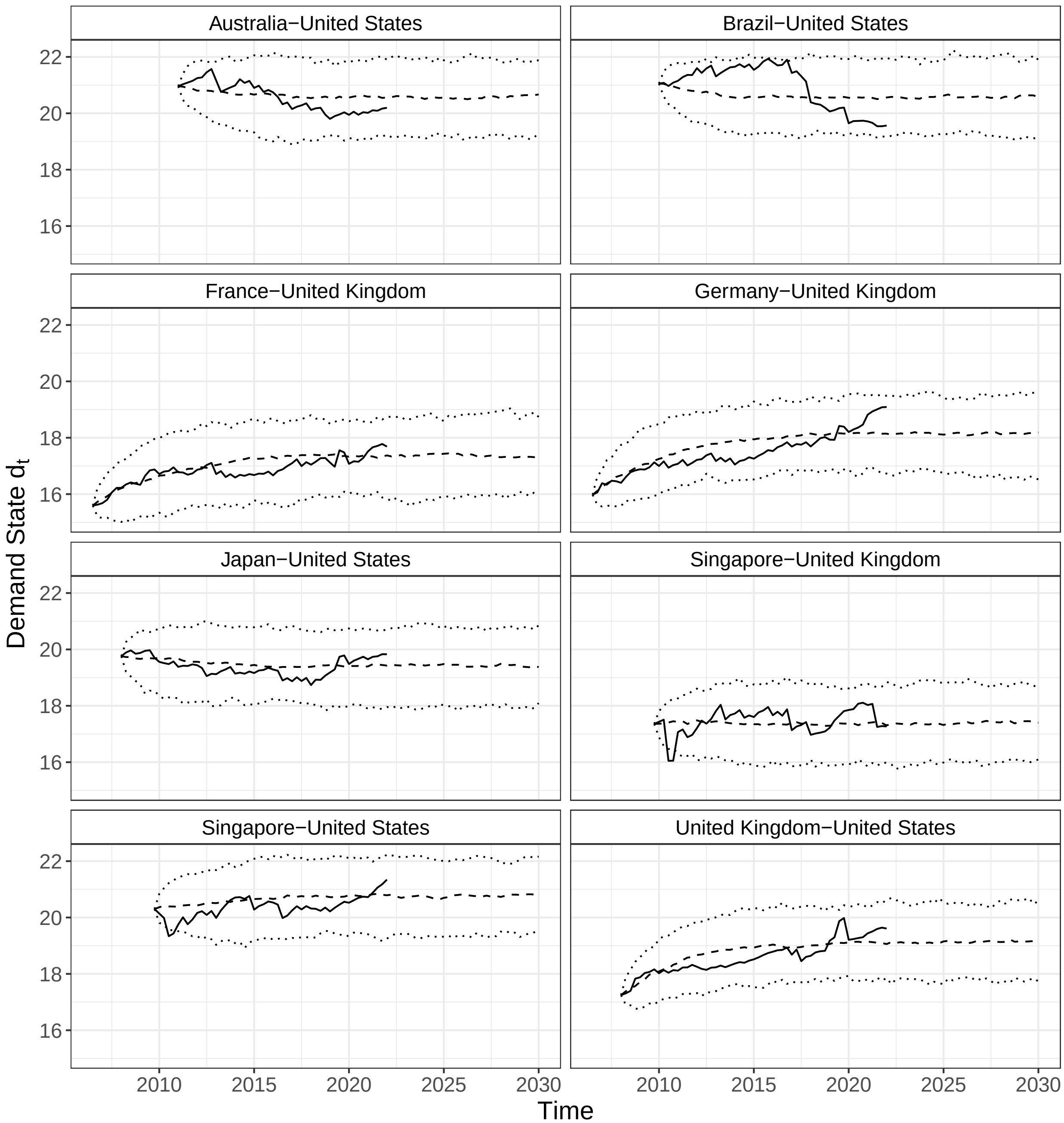
Notes: These figures plot simulated demand curves with the demand for diversification. For illustration purposes, except where otherwise indicated, the parameters used are: $a = 10$, $b = 0.125$, $\mu = 1$, $\sigma^2 = 0.5$, $n = 1$. For further details of the model, see Section 2 of the text.

Table A5: Cable Ownership: Largest Global Firms

Firm Name	Number of Cables
Orange	26
Telecom Italia Sparkle	24
AT&T	23
Tata Communications	21
Arelion (formerly Telia Carrier)	19
C&W Networks	18
Vodafone	18
BT	17
Telkom Indonesia	16
Telstra	16
Verizon	16
Telekom Malaysia	14
XL Axiata	14
China Telecom	13
Singtel	13

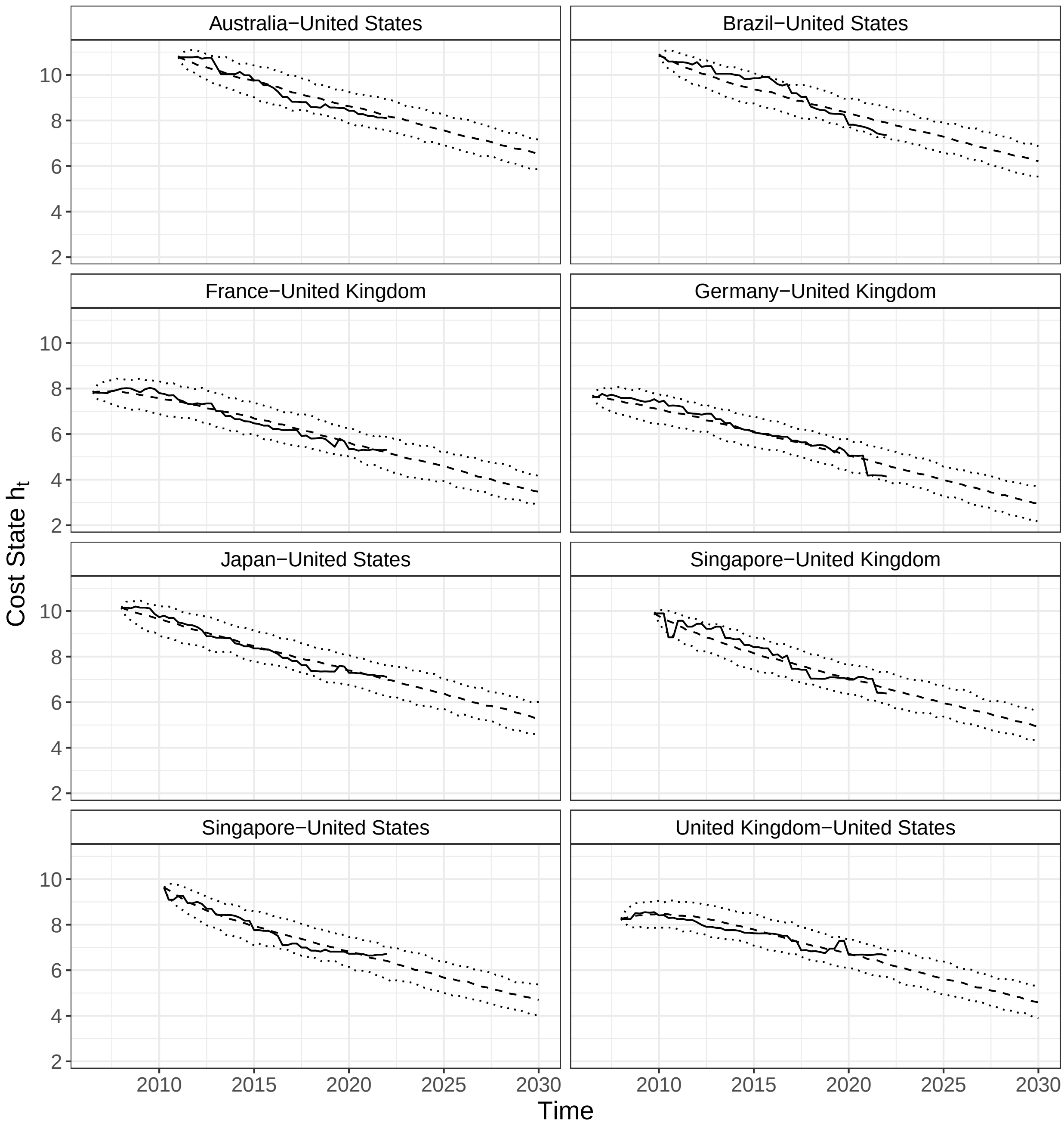
Note: This table shows the largest 15 owners by cable-count in the TeleGeography data. The number of cables corresponds to cables where the owner has a strictly positive ownership share (which is not necessarily 100%).

Figure A5: Demand state over time for a sample of markets



Notes: The figure shows the evolution of d_{mt} over time for a sample of markets (solid line). The dotted lines give the 95% confidence interval from simulations using the transition processes. The dashed line shows the median simulated value.

Figure A6: Cost state over time for a sample of markets



Notes: The figure shows the evolution of h_{mt} over time for a sample of markets (solid line). The dotted lines give the 95% confidence interval from simulations using the transition processes. The dashed line shows the median simulated value.

Table A6: Descriptive Statistics on Cable Faults

Year	Cables	Faults	Propensity (%)
2013	301	7	2.3
2014	316	8	2.5
2015	325	7	2.2
2016	341	7	2.1
2017	356	24	6.7
2018	374	17	4.5
2019	397	15	3.8
2020	418	42	10.0
2021	435	27	6.2
2022	459	14	3.1

Note: Columns 2 and 3 show the number of active cables and faults reported each year. Column 4 shows the annual propensity for a cable to suffer a fault (in percentage terms).

Table A7: Estimation Results: Used Bandwidth (in log) for Top Routes

	OLS			IV	
	(1)	(2)	(3)	(4) First-stage	(5) Second-stage
Bandwidth Price (10G, log)	-0.988***	-0.764***	-0.408***		-1.108***
	(0.0630)	(0.0921)	(0.0629)		(0.260)
Number of undersea cables					
One cable	1.872	1.569	0.111	0.0484	0.161*
	(0.966)	(1.254)	(0.0925)	(0.0775)	(0.0744)
Two cables	1.882*	1.971	1.090*	-0.276***	0.913*
	(0.918)	(1.317)	(0.460)	(0.0815)	(0.419)
Three cables	2.265*	1.802	1.383**	-0.489***	1.058*
	(0.924)	(1.307)	(0.470)	(0.103)	(0.447)
Four cables	3.397***	2.741*	1.538**	-0.540**	1.183*
	(0.907)	(1.323)	(0.534)	(0.168)	(0.487)
Five cables	2.328*	2.017	1.790**	-0.686**	1.336*
	(0.937)	(1.319)	(0.595)	(0.228)	(0.563)
Six cables	1.905*	2.167	1.836**	-0.654**	1.409*
	(0.932)	(1.344)	(0.629)	(0.247)	(0.592)
Seven cables	3.396***	2.554	2.095**	-0.845***	1.538*
	(0.921)	(1.360)	(0.651)	(0.249)	(0.637)
Eight or more cables	3.036**	2.987*	2.425***	-0.917***	1.813**
	(0.989)	(1.313)	(0.581)	(0.226)	(0.584)
Demand factors					
Fixed Broadband Subscriptions (log)		0.392	0.575*	0.303	0.784**
		(0.320)	(0.264)	(0.233)	(0.286)
GDP (log)		-0.273	1.083**	0.283	1.311***
		(0.407)	(0.332)	(0.244)	(0.231)
Aggregate trade flow (log)		0.00663	0.931*	0.222	1.069**
		(0.212)	(0.408)	(0.162)	(0.388)
Number of data/cloud centers (log)		0.574***	0.269**	-0.00361	0.268***
		(0.103)	(0.0778)	(0.0508)	(0.0595)
Distance (km, log)		-0.0964			
		(0.232)			
Common official language		-0.200			
		(0.313)			
Contiguous		0.222			
		(0.408)			
Electricity price (log)				-0.0236	
				(0.0176)	
% change in Electricity price				-0.162***	
				(0.0303)	
Country Pair FEs	No	No	Yes	Yes	Yes
Region Pair × Year FEs	No	No	Yes	Yes	Yes
Weak Identification test					18.52
Endogeneity test (p-value)					11.5 (0.00069)
R^2	0.63	0.75	0.98		0.98
Adjusted R^2	0.63	0.74	0.98		0.98
Observations	1747	1586	1581	1586	1586

Note: The unit of observation is a country pair by quarter. Results are shown for the top ten routes connecting to each region. Standard errors are clustered at the country pair level. Distance corresponds to the bilateral distances between countries, calculated as a weighted arithmetic average of the geodesic distances between the main cities in these countries, where population weights are used. For unilateral variables (GDP, fixed broadband subscriptions, data centers, electricity prices), we restrict the coefficient to be the same for both countries in a pair.

Table A8: Robustness Checks: Used Bandwidth for All Routes

	OLS	IV		IV		
	(1)	(2) 1st-Stage: Price	(3) 2nd-Stage	(4) 1st-Stage: Price	(5) 1st-Stage: Cables	(6) 2nd-Stage
Bandwidth Price (10G, log)	-0.203***		-1.442***			-1.061***
	(0.0538)		(0.277)			(0.262)
Number of undersea cables	0.221***	-0.0811**	0.122*			0.170*
	(0.0448)	(0.0261)	(0.0598)			(0.0793)
Demand factors						
Fixed Broadband Subscriptions (log)	0.177	0.200**	0.429***	0.169*	0.0158	0.271
	(0.107)	(0.0749)	(0.129)	(0.0751)	(0.0450)	(0.146)
GDP (log)	1.112***	0.0949	1.245***	0.202	0.0498	1.575***
	(0.311)	(0.175)	(0.319)	(0.176)	(0.0602)	(0.430)
Aggregate trade flow (log)	-0.0250	0.0293	0.00881	0.0181	-0.00626	-0.00518
	(0.0513)	(0.0364)	(0.0844)	(0.0273)	(0.00652)	(0.0593)
Number of data/cloud centers (log)	0.111	0.0195	0.143	-0.00584	-0.0486***	0.125
	(0.0835)	(0.0638)	(0.118)	(0.0865)	(0.0143)	(0.131)
Electricity price (log)		0.0496*		0.0867***	-0.0253*	
		(0.0217)		(0.0251)	(0.00997)	
% change in Electricity price		-0.224***		-0.256***	-0.0144	
		(0.0367)		(0.0372)	(0.0204)	
Number of cable faults in $t-1$				-0.0246*	-0.0191**	
				(0.0117)	(0.00595)	
Number of cable faults in $t-2$				-0.0000986	0.0109*	
				(0.0110)	(0.00532)	
Number of undersea cables in $t-1$				-0.0665	0.933***	
				(0.0369)	(0.0144)	
Country Pair FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region Pair × Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Weak Identification test			21.02			11.86
Endogeneity test (p-value)			31.0 (0)			35.5 (0)
R^2	0.97		0.95			0.97
Adjusted R^2	0.97		0.95			0.97
Observations	3849	3863	3863	2510	2510	2510

Note: The unit of observation is a country pair by quarter. Results are shown for the top ten routes connecting to each region. Standard errors are clustered at the country pair level. Distance corresponds to the bilateral distances between countries, calculated as a weighted arithmetic average of the geodesic distances between the main cities in these countries, where population weights are used. For unilateral variables (GDP, fixed broadband subscriptions, data centers, electricity prices), we restrict the coefficient to be the same for both countries in a pair.

Table A9: Fit of equilibrium quantity-setting equation

	Dependent variable: Price	
	Estimate	S.E.
Markup	1.285	(0.226)
Electricity Price	170.119	(69.520)
% Change in Electricity Price	-415.521	(1172.760)
Number of cable faults	167.727	(422.788)
Cable length	-0.057	(0.085)
R^2	0.756	
Adjusted R^2	0.755	
Observations	2731	

Note: The unit of observation is the market (country pair) by quarter. Standard errors are clustererd at the market level and shown in parenthesis. All variables are in level. Cable length is averaged over all cables operating in a given market.