Investment and Usage of the Subsea Internet Cable Network

Jihye Jeon
Boston University

Marc Rysman
Boston University

December 10, 2025

Subsea Internet Cable Network

- Critical infrastructure for global communication, cloud services, and AI.
- More than 99% of overseas data traffic travels through subsea cables.
- The global internet traffic grew by a factor of 1000 between 2002 and 2020.
- As of 2021, there are over 400 active subsea cables spanning 1.5 million kilometers in length, connecting 182 countries.

Clark (2018) on fiber optic cable:

"The economic implications of its development and deployment are perhaps the single most important factor in the growth and success of the Internet."

Jayne Stowell, Google (2019):

"People think that data is in the cloud, but it's not. It's in the ocean."

Potential Inefficiencies in Subsea Cable Network Investment

1. Network nature of the market
   - An investment in one region affects data flows in the rest of the world.
   - A cable between North America and Europe may carry data going from Africa to South America
2. Business stealing incentives
   - A new cable between North America and Europe steals share from existing cables, and possibly from other markets.

Potential Inefficiencies in Subsea Cable Network Investment

1. Network nature of the market
   - An investment in one region affects data flows in the rest of the world.
   - A cable between North America and Europe may carry data going from Africa to South America
2. Business stealing incentives
   - A new cable between North America and Europe steals share from existing cables, and possibly from other markets.

Research question

- How much does cable construction contribute to growth of usage?
- How does private investment in cables deviate from the social optimum?
- What factors determine misallocation?

Our Approach

- We utilize a new data set with the history of cable construction and usage.
- We develop a structural model international data flows.
  - Country-to-country demand for data exchange.
  - How data traverses the cable network.
- We develop an equilibrium model of cable construction that addresses:
  - The networked nature of the market.
  - The possibility of joint investments by consortiums of firms.
  - Differing incentives of content providers
- We use our results to contrast social and private payoffs in cable construction.

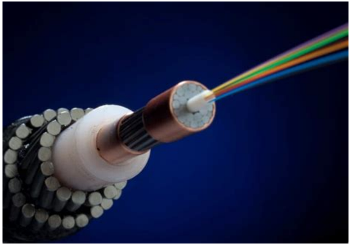
Literature Review

- Subsea cables:
  - Hjort & Poulsen (2019) shows that the arrival of a cable in Africa contributes to development.
  - Caoui & Steck (2023) study the role of path diversity in this market.
- Infrastructure and trade:
  - Jeon (2022), Brancaccio, Kalouptsidi, & Papgeorgiou (2020), Frechette, Lizzeri, & Salz (2019), Allen & Arkolakis (2023).
- Networks retail, banking, and airlines:
  - Jia (2008), Holmes (2011); Houde, Newberry, & Seim (2022), Aguiregabiria & Ho (2010); Li, Mazur, Park, Roberts, Sweeting & Zhang.
- Internet services and telecom investment
  - Greenstein (2015), Nevo, Turner, & Williams (2016), Rysman (2016), Elliot, Houngbonon, Ivaldi, & Scott (2023), Bourreau, Grzybowski & Muñoz-Acevedo (2023), Granja (2002)

Investment and Distance Matter

- Smith (2009) and Leighton (2009) discuss investment and particularly distance in affecting internet quality.
- Clark & Wedeman (2024) provide new measures of internet quality and emphasize distance.
- Fanou, Huffaker, Bradley, Mok & claffy (2020) show measurable improvement of internet speed and packet loss in Africa and South America after new cable between them.
- Malone, Nevo, & Williams (2020) and Elliott, Houngbonon, Ivaldi & Scott (2023) emphasize consumer responsiveness to telecom investment in Internet quality.

Cables!



Industry

Cable: Cable owners employ a management firm to sell access

↓

Internet Backbone: e.g. Level3

↓

Access network: e.g. Verizon AT&T Comcast

Industry

Investors are changing

- Traditionally, cables have were owned by national carriers.
  - Examples: AT&T, BT, NTT, or China Telecom.
- Recently, content providers are major investors.
  - Google, Microsoft, Amazon, Meta.

Level of economic regulation is very low.

- In the US, FCC must approve all landings.
- It primarily considers national security issues, such as:
  1. Robustness of landing to abuse.
  2. Equipment from untrustworthy firms.
  3. Other landings in problematic countries.

Data Source

- We have data from Telegeography, a market research firm that covers internet equipment markets with a special focus on subsea cables.
- Annual data from 2002-2020.
- We observe cable features.
  - Construction date, decommission date, landing points
  - Potential capacity, length.
  - Owners, construction cost.
- We observe bandwidth usage at the region-pair level.
  - i.e., Bandwidth usage between North America and Europe.
  - Dependent variable!

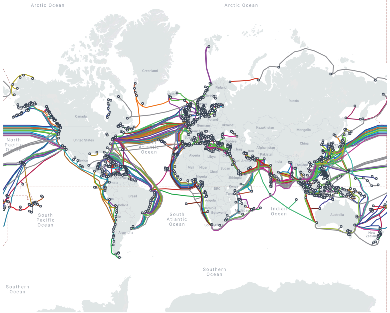
Data Source

- We have data from Telegeography, a market research firm that covers internet equipment markets with a special focus on subsea cables.
- Annual data from 2002-2020.
- We observe cable features.
  - Construction date, decommission date, landing points
  - Potential capacity, length.
  - Owners, construction cost.
- We observe bandwidth usage at the region-pair level.
  - i.e., Bandwidth usage between North America and Europe.
  - Dependent variable!
- CEPII data on international trade.

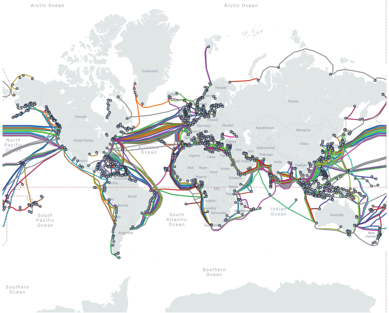
Data Issues

- We observe bandwidth usage at region-pair level but capacity at cable level.
  - Aggregate capacity between regions to get a single region-pair number.
- We observe bandwidth usage between two regions, not the endpoints in the data flows.
  - Build a model that predicts variable we observe.
- We do not observe terrestrial cables.
  - Assume terrestrial connectivity is uniform within regions.

Aggregating Cable Data



Aggregating Cable Data



- We sum up capacity over all cables with landing points in two regions to find capacity for that market.
- We consider all paths between regions with no more than two stops.

Observations

Demand

- 399 market-year observations.
- 133 island-year observations.
- 179,778 country-pair-year observations.

Observations

Demand

- 399 market-year observations.
- 133 island-year observations.
- 179,778 country-pair-year observations.

Investment

- 102 cables.
- 197 owners.
- 2,798 owner-year observations.

- Annual data from 2002-2020.

Definitions

- **Countries** are partitioned into **regions**.
  - Regions are groups of countries such that traffic between regions must traverse the ocean.
  - North America, South America, Sub-Saharan Africa, Oceania, East Asia, South Asia, Europe (Middle East, North Africa).
- **Markets** are pairs of regions.
  - Regions are connected by cables.
  - All countries in the same regions have access to the same connectivity.
  - Example: Europe-North America
- **Paths** are sets of markets that connect any two regions.
  - Example: Paths between the UK and US include:
    - Direct: Europe-North America.
    - Indirect: Europe-East Asia-North America, Europe-South America-North America.

Model: Simultaneous Discrete Game

- Firms choose simultaneously whether or not to invest in each market.
- Demand is realized.
  - Country-to-country quantity traverses cable network.
  - Firms profit from quantity on the part of the network that they own.
  - Investment in one market affects worldwide traffic flows.

Cable Investment

- In each market $m = 1, \ldots, M$, an opportunity to construct a cable arises.
- Firms $j = 1, \ldots, J$ simultaneously make investment choice $a_{jmt} \in \{0, 1\}$.
- Denote choices in market m as $\boldsymbol{a}_{mt} = (a_{jmt})_j$ and all choices as $\boldsymbol{A}_t = (\boldsymbol{a}_{mt})_m$.
- Added capacity is $\tilde{K}_{mt}(\boldsymbol{a}_{mt})$.
  - Known deterministic function.
- Implies cost $C_{mt}(\tilde{K}_{mt}(\boldsymbol{a}_{mt}))$.
- Multiple firms may choose to build a given opportunity.
  - They become joint owners, evenly splitting ownership and the cost.

Firm's Ownership and Cost

- Firm j's added capacity $\tilde{k}_{jmt}(\boldsymbol{a}_{mt})$ and cost is:

$$\tilde{k}_{jmt}(\boldsymbol{a}_{mt}) = \frac{a_{jmt}}{\sum_{z=1}^{J} a_{zmt}} \tilde{K}_{mt}\left(\boldsymbol{a}_{mt}\right) \qquad c_{jmt}\left(\boldsymbol{a}_{mt}\right) = \frac{a_{jmt}}{\sum_{z=1}^{J} a_{zmt}} C_{mt}\left(\tilde{K}_{mt}\left(\boldsymbol{a}_{mt}\right)\right).$$

Firm's Ownership and Cost

- Firm j's added capacity $\tilde{k}_{jmt}(\boldsymbol{a}_{mt})$ and cost is:

$$\tilde{k}_{jmt}(\boldsymbol{a}_{mt}) = \frac{a_{jmt}}{\sum_{z=1}^{J} a_{zmt}} \tilde{K}_{mt}(\boldsymbol{a}_{mt}) \qquad c_{jmt}(\boldsymbol{a}_{mt}) = \frac{a_{jmt}}{\sum_{z=1}^{J} a_{zmt}} C_{mt}\left(\tilde{K}_{mt}(\boldsymbol{a}_{mt})\right).$$

- Market and firm capacity evolves:

$$K_{mt}(\boldsymbol{a}_{mt}, K_{mt-1}) = K_{mt-1} + \tilde{K}_{mt}(\boldsymbol{a}_{mt}) \quad k_{jmt}(\boldsymbol{a}_{mt}, k_{jmt-1}) = k_{jmt-1} + \tilde{k}_{jmt}(\boldsymbol{a}_{mt}).$$

Firm's Ownership and Cost

- Firm j's added capacity $\tilde{k}_{jmt}(\boldsymbol{a}_{mt})$ and cost is:

$$\tilde{k}_{jmt}(\boldsymbol{a}_{mt}) = \frac{a_{jmt}}{\sum_{z=1}^{J} a_{zmt}} \tilde{K}_{mt}\left(\boldsymbol{a}_{mt}\right) \qquad c_{jmt}\left(\boldsymbol{a}_{mt}\right) = \frac{a_{jmt}}{\sum_{z=1}^{J} a_{zmt}} C_{mt}\left(\tilde{K}_{mt}\left(\boldsymbol{a}_{mt}\right)\right).$$

- Market and firm capacity evolves:

$$K_{mt}(\boldsymbol{a}_{mt}, K_{mt-1}) = K_{mt-1} + \tilde{K}_{mt}(\boldsymbol{a}_{mt}) \quad k_{jmt}(\boldsymbol{a}_{mt}, k_{jmt-1}) = k_{jmt-1} + \tilde{k}_{jmt}(\boldsymbol{a}_{mt}).$$

- Number of cables evolves:

$$n_{mt}(\boldsymbol{a}_{mt}, n_{mt-1}) = n_{mt-1} + \mathbb{1}\{a_{jmt} = 1 \text{ for any } j = 1, \ldots, J\}$$

Firm Profits

Firm j's expected profit is

$$\pi_{jt}\left(\boldsymbol{A}_t, \mathcal{S}_t\right) = E\left[\sum_{m \in \mathcal{M}} \left\{ \underbrace{r_{jmt}\left(\boldsymbol{A}_t, \mathcal{S}_t; \gamma\right)}_{\text{revenue}} - \underbrace{c_{jmt}\left(\boldsymbol{a}_{mt}\right)}_{\text{investment cost}} + \underbrace{\nu_{ajmt}}_{\text{cost shock}} \right\} \middle| \mathcal{J}_{jt} \right].$$

- $\boldsymbol{A}_t$ contains a_{jmt} for all jm.
- $\mathcal{S}_t$ contains firm capacities in every market.
- $\mathcal{J}_{jt}$ is the firm's information set.

Firm Profits

Firm j's expected profit is

$$\pi_{jt}\left(\boldsymbol{A}_t, \mathcal{S}_t\right) = E\left[\sum_{m\in\mathcal{M}}\left\{\underbrace{r_{jmt}\left(\boldsymbol{A}_t, \mathcal{S}_t; \gamma\right)}_{\text{revenue}} - \underbrace{c_{jmt}\left(\boldsymbol{a}_{mt}\right)}_{\text{investment cost}} + \underbrace{\nu_{ajmt}}_{\text{cost shock}}\right\}\middle|\mathcal{J}_{jt}\right].$$

- $\boldsymbol{A}_t$ contains a_{jmt} for all jm.
- $\mathcal{S}_t$ contains firm capacities in every market.
- $\mathcal{J}_{jt}$ is the firm's information set.

Firms play a Nash equilibrium, simultaneously maximizing their profits.

$$\text{Firm } j \text{ solves } \max_{a_{jmt},\, m=1,\dots,M} \pi_{jt}\left(\boldsymbol{A}_t, \mathcal{S}_t\right).$$

Revenue

- We specify the variable-profit function as:

$$r_{jmt}(\boldsymbol{A}_t, \mathcal{S}_t) = \gamma\, \underbrace{q_{jmt}(\boldsymbol{A}_t, \mathcal{S}_t)}_{\text{quantity}} =$$

Revenue

- We specify the variable-profit function as:

$$r_{jmt}(\boldsymbol{A}_t, \mathcal{S}_t) = \gamma \underbrace{q_{jmt}(\boldsymbol{A}_t, \mathcal{S}_t)}_{\text{quantity}} = \gamma \underbrace{D_{mt}(\boldsymbol{A}_t, \mathcal{S}_t)}_{\text{total data}} \underbrace{\frac{k_{jmt}\left(\boldsymbol{a}_{mt}\right)}{K_{mt}\left(\boldsymbol{a}_{mt}\right)}}_{\text{ownership share - market}} .$$

- $D_{mt}(\boldsymbol{A}_t, \mathcal{S}_t)$ depends on global demand and cable capacity.
- Estimate γ with moment inequalities framework.

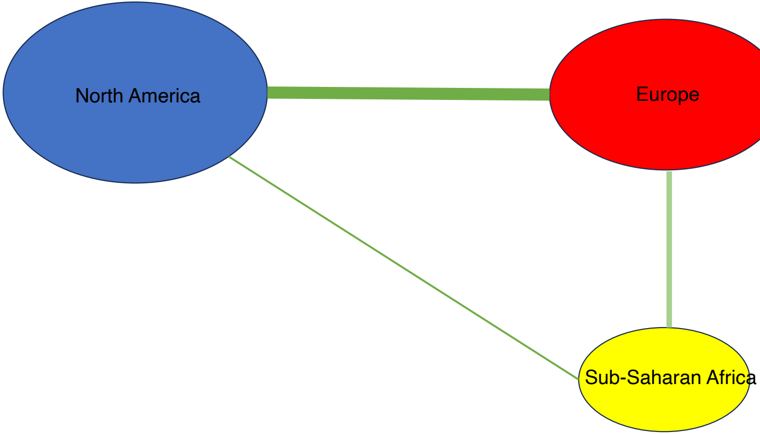
Model: Data Flows

- A pair of countries c and k has a potential demand for data of $\overline{M}$.
- The total data transmitted between countries c and k at time t

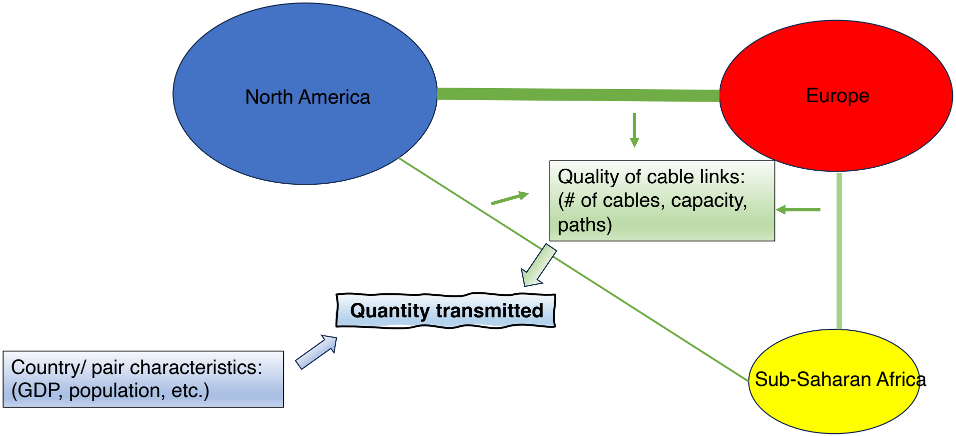
$$d_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \frac{\exp(x_{ckt}\theta^d + v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t))}{1 + \exp\left(x_{ckt}\theta^d + v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t)\right)}\overline{M}$$

  - x_{ckt} includes time trends and country and country-pair characteristics.
  - $v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t)$ is the quality of the connection between the two countries.
- Outside option captures that some data is lost or never searched.

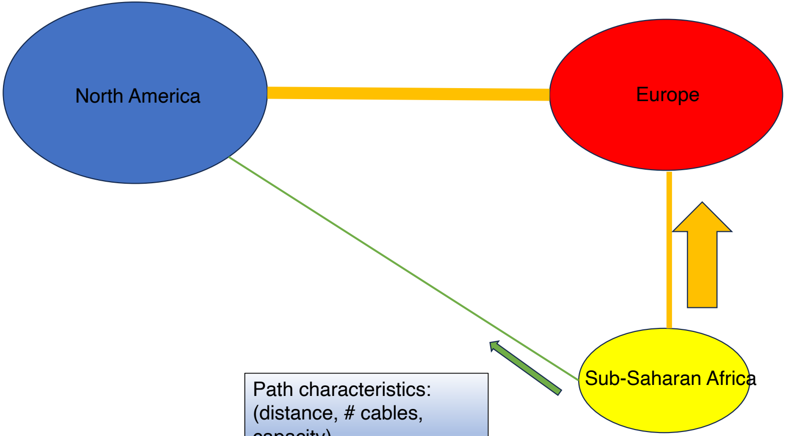
Data Flows



Data Flows


North America
Europe
Quality of cable links:
(# of cables, capacity,
paths)
Quantity transmitted
Country/ pair characteristics:
(GDP, population, etc.)
Sub-Saharan Africa

Allocation Across Paths



Path characteristics:
(distance, # cables, capacity)

Model: Cable Usage

- Data between two countries get allocated across many potential paths.
- Country pair ck has P_{ckt} *paths* indexed by $p = 1, \ldots, P_{ckt}$.
- Specify a reduced-form process for how data gets allocated to paths.

Model: Cable Usage

- Data between two countries get allocated across many potential paths.
- Country pair *ck* has P_{ckt} *paths* indexed by $p = 1, \ldots, P_{ckt}$.
- Specify a reduced-form process for how data gets allocated to paths.
- Let $\delta_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t)$ be the attractiveness of the path.
- The share of data going between c and k on path p at time t is:

$$s_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \frac{\exp\left(\delta_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t)\right)}{\sum_{z=1}^{P_{ckt}} \exp\left(\delta_{zckt}(\boldsymbol{A}_t, \mathcal{S}_t)\right)}$$

Model: Path Attractiveness

- Path attractiveness defined as:

$$\delta_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \theta^{\delta,0} + \theta^{\delta,d} \ln(\mathtt{dist}_{pck}) + \theta^{\delta,K} K^{\delta}_{pckt}(\boldsymbol{A}_{mt}, \mathcal{S}_t) + \theta^{\delta,n} n^{\delta}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t).$$

- $\mathtt{dist}_{pck}$ is path length.
- $K^{\delta}_{pckt}(\boldsymbol{A}_{mt}, \mathcal{S}_t)$ is path capacity.
- $n^{\delta}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t)$ is number of cables on path.
  - Functions of features of underlying markets.
  - $K^{\delta}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \min_{m \in \mathcal{M}_{pckt}} K_{mt}(\boldsymbol{a}_{mt}, K_{mt-1})$
  - $n^{\delta}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t)$ is capacity weighted average of $n_{mt}(\boldsymbol{a}_{mt-1}, n_{mt-1})$.

Cable Quality Between Countries

- We let $v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t)$ equal the inclusive value of logit choice among cables:

$$v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \ln \left(\sum_{p=1}^{P_{ckt}} \exp\left(\delta_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t)\right) \right).$$

Cable Quality Between Countries

- We let $v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t)$ equal the inclusive value of logit choice among cables:

$$v_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \ln \left(\sum_{p=1}^{P_{ckt}} \exp\left(\delta_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t)\right) \right).$$

- The amount of data traveling on path *pckt* is:

$$\hat{d}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) = s_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) d_{ckt}(\boldsymbol{A}_t, \mathcal{S}_t).$$

Adding Up

- Total quantity between c and k in market m:

$$\tilde{d}_{mckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \sum_{p=1}^{P_{ckt}} \hat{d}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) \mathbb{1}\{m \in \mathcal{M}_{pckt}\}.$$

  - $\mathcal{M}_{pckt}$ is the set of markets on path p between country pair ck.

Adding Up

- Total quantity between c and k in market m:

$$\tilde{d}_{mckt}(\boldsymbol{A}_t, \mathcal{S}_t) = \sum_{p=1}^{P_{ckt}} \hat{d}_{pckt}(\boldsymbol{A}_t, \mathcal{S}_t) \mathbb{1}\{m \in \mathcal{M}_{pckt}\}.$$

  - $\mathcal{M}_{pckt}$ is the set of markets on path p between country pair ck.
- Add up to market-level quantity and region-level quantity:

$$D_{mt}(\boldsymbol{A}_t, \mathcal{S}_t) = \sum_{c=1}^{C} \sum_{k=1}^{c-1} \tilde{d}_{mckt}(\boldsymbol{A}_t, \mathcal{S}_t).$$

- Perform method of moments.

Estimation of Demand: GMM

- Error terms:

$$\xi_{mt}(\theta) = D^*_{mt} - D_{mt}(\theta) \quad \Delta\xi_{mt}(\theta) = \xi_{mt}(\theta) - \xi_{mt-1}(\theta).$$

- Moments:

$$\boldsymbol{m}(\theta) = \boldsymbol{Z}'\xi(\theta) \quad \boldsymbol{m}^{\Delta}(\theta) = \boldsymbol{Z}^{\Delta\prime}\boldsymbol{\Delta}\xi(\theta).$$

- Instruments: Demographics, time
- Capacity may be endogenous to usage.
- Include lagged value of capacity in $\boldsymbol{Z}^{\Delta}$.
  - Firms can't anticipate changes in usage beyond time trend.

Estimation of Supply: Moment Inequalities

- For investors:

$$E\left[\sum_{z\in\mathcal{M}} \gamma_{jt} D_{mt}(\boldsymbol{A}_t, \mathcal{S}_t)\frac{k_{jmt}(\boldsymbol{a}_{mt})}{K_{jmt}(\boldsymbol{a}_{mt})} - c_{jmt}(\boldsymbol{a}_{jmt}) + \Delta\nu_{jmt} \middle| a_{jmt} = 1, \mathcal{J}_{jt}\right] \geq 0.$$

  - Analogous for non-investors.
- Eliminate structural error $\Delta\nu_{jmt}$: Difference across similar investors and non-investors in the same market.
  - Assumes $\Delta\nu_{jmt} = \Delta\nu_{j'mt}$ for similar firms j,j'.
  - Follows Pakes, Porter, Ho, Ishii (2015).

Estimation of Supply: Dynamics

A robustness check

- Assumption of static decision making is obviously problematic.
- Typical problem is how to specify how rivals respond to deviations in future.

Assume firms have perfect foresight and commitment.

Firms play Nash equilibrium!

Estimation of Supply: Dynamics

A robustness check

- Assumption of static decision making is obviously problematic.
- Typical problem is how to specify how rivals respond to deviations in future.

Assume firms have perfect foresight and commitment.

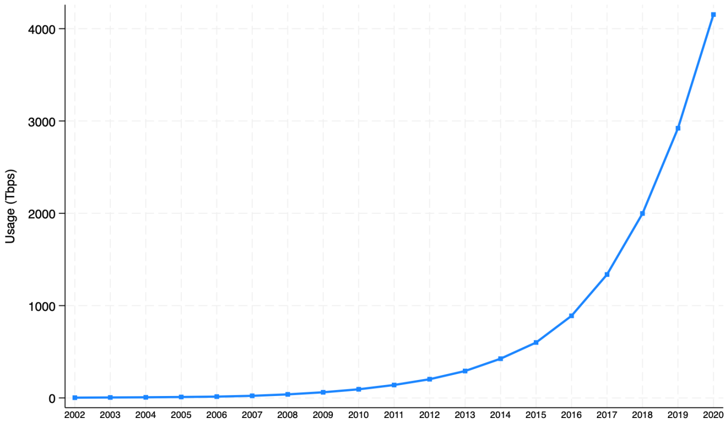
Firms play Nash equilibrium!

- Estimate with single deviation moment inequalities.
- Rivals do not adjust to deviations.
- Sum profits over all years.

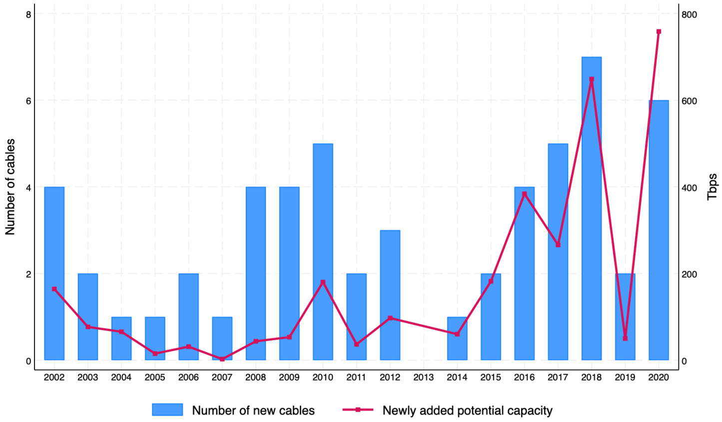
Dynamics: Implementation

- Break up into three time periods (groups of years).
- Estimate separately for each time period.
- Firms make all of their choices in first year of period.
- Assume states will say the same after last year of period.
- Estimate with single-deviation moment inequalities summing profit over future years.
  - Assume rivals do not adjust.

Bandwidth Usage from 2002-2020

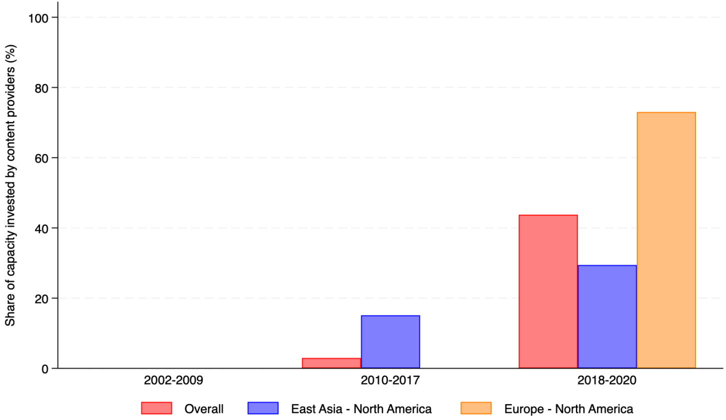


Investment



- 58 new cables and 3,185 Tbps of new capacity were added from 2002 to 2020.

Increased Role of Content Providers



- In 2018-2020, content provider investments accounted for 42% of the total new capacity.

Demand Estimates

Panel A. Demand for data (x_{ckt})	θ^d	
Constant	-182.32	(71.33)
GDP	1.26	(0.12)
Population	-0.30	(0.11)
Time trend	0.36	(0.01)
2010-2017	-0.47	(0.03)
2018-2020	-0.80	(0.06)
Market fixed effects	Y	
Panel B. Cable usage (Z_{pckt})	θ^δ	
Constant	126.15	(70.21)
Path length	-1.82	(0.09)
Potential capacity	1.16	(0.04)
Number of cables	0.12	(0.01)

Demand Estimates

Panel A. Demand for data (x_{ckt})	θ^d	
Constant	-182.32	(71.33)
GDP	1.26	(0.12)
Population	-0.30	(0.11)
Time trend	0.36	(0.01)
2010-2017	-0.47	(0.03)
2018-2020	-0.80	(0.06)
Market fixed effects	Y	
Panel B. Cable usage (Z_{pckt})	θ^δ	
Constant	126.15	(70.21)
Path length	-1.82	(0.09)
Potential capacity	1.16	(0.04)
Number of cables	0.12	(0.01)

- A 1% increase in GDP would increase data flows by 1.2%.
- A 1% increase in population would decrease data flows by 0.3%.
- Other specifications include:
  - Common language, GDPR, trade agreements, number of internet subscribers, etc.

Demand Estimates

Panel A. Demand for data (x_{ckt})	θ^d	
Constant	-182.32	(71.33)
GDP	1.26	(0.12)
Population	-0.30	(0.11)
Time trend	0.36	(0.01)
2010-2017	-0.47	(0.03)
2018-2020	-0.80	(0.06)
Market fixed effects	Y	
Panel B. Cable usage (Z_{pckt})	θ^δ	
Constant	126.15	(70.21)
Path length	-1.82	(0.09)
Potential capacity	1.16	(0.04)
Number of cables	0.12	(0.01)

- A 1% increase in GDP would increase data flows by 1.2%.
- A 1% increase in population would decrease data flows by 0.3%.
- Other specifications include:
  - Common language, GDPR, trade agreements, number of internet subscribers, etc.
- A shorter, higher capacity path with a larger number of cables is preferred.

Cost and Capacity Regressions

	ln(capacity)	ln(cost)
Time	0.144***	-0.028***
	(0.027)	(0.011)
ln(capacity)		0.137***
		(0.044)
ln(length)		0.941***
		(0.045)
Constant	7.764***	9.845***
	(0.993)	(0.520)
Region FEs	Yes	No
Observations	71	68
R-squared	0.553	0.847

- Capacity increases over time and costs fall.
- Total effect of time on cost is slightly negative.
- Costs are more sensitive to distance than capacity.

Supply Model Estimates

95% confidence set for the profit parameter γ_{jt} ($M/ Tbps)

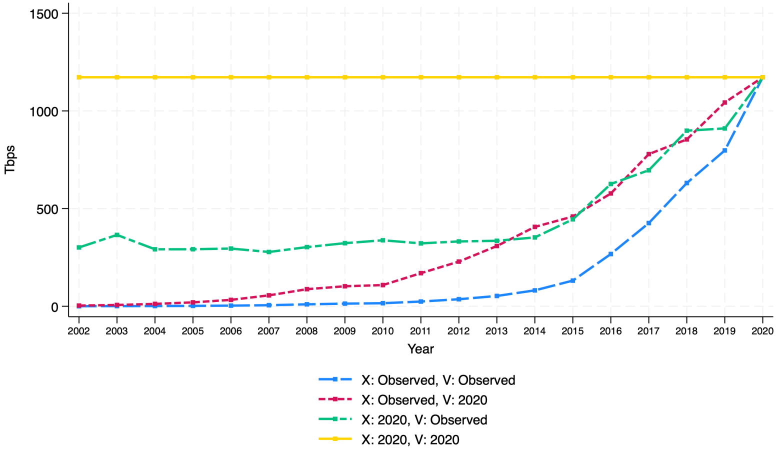
	Carriers	Content providers
2002-2009	[2674.3 , 3517.9]	NA
2010-2017	[22.3 , 98.8]	[11.3 , 96.0]
2018-2020	[12.0 , 15.6]	[5.0 , 18.8]

- A 1 Tbps increase in a firm's quantity would lead to a profit increase of $2.7-3.5B in 2002-2009.
- Steep decline in line with the decline in bandwidth prices.
- Lower profit for content providers in line with their distinct motives and market position.

Robustness

- No selection on unobservables ($\nu_{jmt} = 0$).
- Alternative differencing strategy.
- Allow markup γ_{jt} to differ across large and small firms.
- Allow counterfactual capacity $\tilde{K}_j(\boldsymbol{a}_{mt})$ to change of one of several firms joins or drops out.
- Use only large owners to form moments.
- Reweighting of markets relative to firms.
- Dynamic model: Full commitment, perfect foresight.

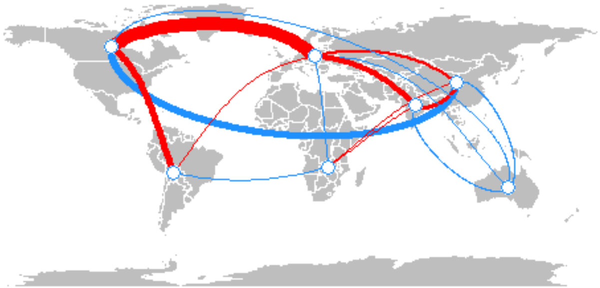
Decomposition of Growth in Data Flows: Demographics vs. Cable Network



- In 2002-2012, endowing the world with the 2020 cable network (vs. x_{ckt}) has a smaller effect.
- By 2013, the two have comparable effect (a 6x increase).

Counterfactuals: Efficient Network

Where would the social planner build?



- Planner would build new cables in 11 markets (in red) in 2020.
- Entrants would find 7 markets profitable.
  - This matches average number of investments in 2018-20.
- Includes large and small markets.

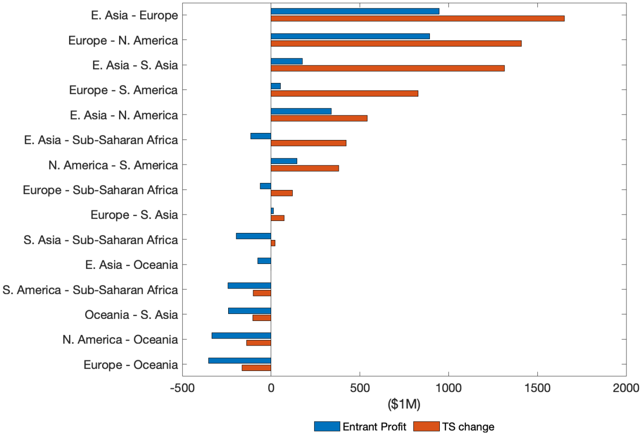
Counterfactuals: Efficient Network

Questions:

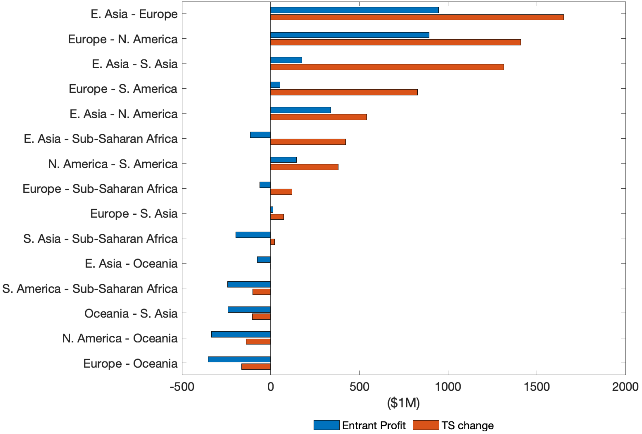
- Why do investments in certain markets generate higher social value?
- Is there misalignment and what are the sources?

- Calculate consumer surplus using "back of the envelope" calculation with CED and markup γ_{jt} to get elasiticy.

Counterfactuals: Private vs. Social Incentives

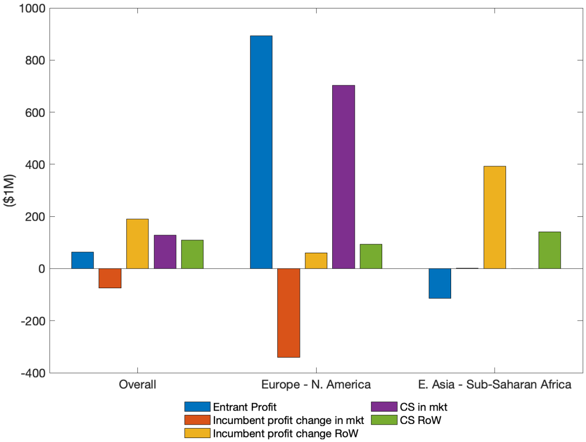


Counterfactuals: Private vs. Social Incentives



- Entrant incentives are not always aligned with what is socially optimal.
- Entrant profit$<$ Total surplus change. $\Rightarrow$ Business stealing not driving misalignment.

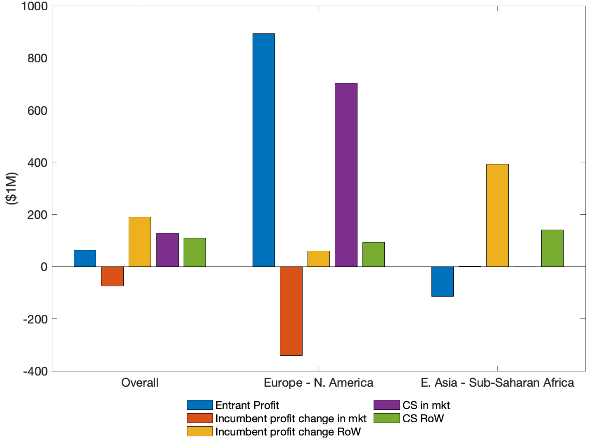
Counterfactuals: Sources of Inefficiencies and Entry Subsidies



EU-N. America

- Large business stealing
- Small out-of-market externalities
- Large social surplus and entrant profit due to strong demand within market

Counterfactuals: Sources of Inefficiencies and Entry Subsidies



E. Asia-SSA & EU-S. Am

- Little business stealing
- Large out-of-market externalities
- Subsidies of \$30-170M would be required to incentivize de novo entrants
- Individual governments would not be sufficiently incentivized to provide these subsidies due to most of the benefits falling outside the market.

Conclusion

- This paper studies investments in subsea internet cables, a critical piece of communication infrastructure.
- We develop and estimate a model of cable usage and investment.
  - Maps country-to-country demand for data int observed data flows.
  - Accounts for global traffic flows and the formation of consortiums of investors.
- We show that the construction of the cable network contributes substantially to the growth of trade in data.
- We find underinvestment relative to social optimum.
- Spillovers to other markets drives result.

Demand-Side Parameters

- We match $D_{mt}(\theta)$ and $Q_{rt}(\theta)$ to market and island-country used bandwidth.

$$\xi^D_{mt}(\theta) = D^*_{mt} - D_{mt}(\theta) \qquad \xi^Q_{rt}(\theta) = Q^*_{rt} - Q_{rt}(\theta)$$
$$\Delta\xi^D_{mt}(\theta) = \xi^D_{mt}(\theta) - \xi^D_{mt-1}(\theta) \qquad \Delta\xi^Q_{rt}(\theta) = \xi^Q_{rt}(\theta) - \xi^Q_{rt-1}(\theta).$$

- Assume instruments $\boldsymbol{y}^D_{mt}$ imply $E[\xi^D_{mt}(\theta^0)|\boldsymbol{y}^D_{mt}] = 0$, etc.
  - at optimal parameters θ^0.
- $\boldsymbol{y}^D_{mt}$ average over markets and time, interact with time.
- Include lagged values K_{mt-1} and n_{mt-1} in $\mathbf{y}^{D,\Delta}_{mt}$.

Moment Inequalities Based on Revealed Preference

- Revealed preference implies that if firm j invests in market m, t:

$$\sum_{z \in \mathcal{M}} \Delta r_{jmzt}\left(\boldsymbol{A}_t, \mathcal{S}_t; \gamma\right) + \Delta \varepsilon_{jmt} - c_{jmt}\left(\boldsymbol{a}_{mt}^{j,1}\right) + \Delta \nu_{jmt} \geq 0$$

  - Δr_{jmzt} is the change in the revenue in market z from investing in mt
  - $\Delta \varepsilon_{jmt} = \varepsilon_{1jmt} - \varepsilon_{0jmt}$, $\Delta \nu_{jmt} = \nu_{1jmt} - \nu_{0jmt}$
  - $\boldsymbol{a}_{mt}^{j,1} = \boldsymbol{a}_{mt}$ with element $j = 1$

Moment Inequalities Based on Revealed Preference

- We make assumptions to derive inequalities that can be applied to the data:
  - $\Delta\nu_{jmt} = \Delta\nu_{j'mt}$ if j and j' have the same firm type for all mt.
  - Firms have rational expectations: $E[\Delta\epsilon_{jmt}|\mathcal{J}_{jt}] = 0$.
  - There are elements of information set observed to the researcher: $Z_{jmt} \subsetneq \mathcal{J}_{jt}$.
- Sum inequalities across pairs with $a_{jmt} = 1, a_{j'mt} = 0$ within firm type, market, year:

$$E\left[\sum_{z\in\mathcal{M}} \Delta r_{jmzt}(\boldsymbol{A}_t, \mathcal{S}_t; \gamma) - \sum_{z\in\mathcal{M}} \Delta r_{j'mzt}(\boldsymbol{A}_t, \mathcal{S}_t; \gamma) - c_{jmt}\left(\boldsymbol{a}_{mt}^{j,1}\right) + c_{j'mt}\left(\boldsymbol{a}_{mt}^{j',1}\right) \middle| a_{jmt} = 1, a_{j'mt} = 0, \mathcal{Z}_{jmt}, \mathcal{Z}_{j'mt}\right] \geq 0.$$

- We have an analogous inequality for the case of no investment.

Estimation: Added capacity under counterfactual investment choices

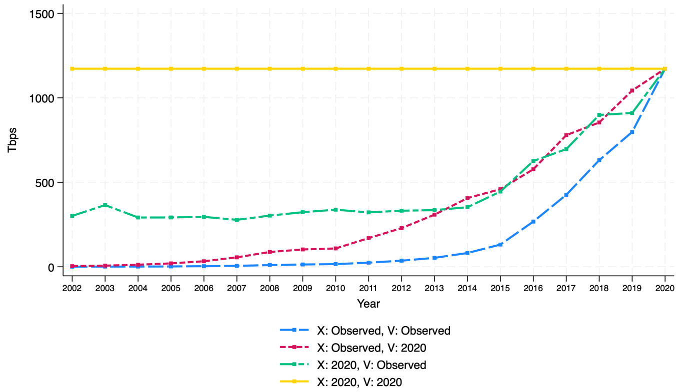
- What is counterfactual capacity level if j deviates to generate $\boldsymbol{a}_{mt}$?

$$\tilde{K}_j(\boldsymbol{a}_{mt}) = \begin{cases} 0 & \text{if } a_{zmt} = 0 \; \forall \; z \\ \widehat{\tilde{K}}_{mt} & \text{if } a_{jmt} = 1 \text{ and } a_{zmt} = 0 \; \forall z \neq j \\ \tilde{K}^{\texttt{data}}_{mt} & \text{if } a_{zmt} = 1 \text{ for any } z \neq j \end{cases} \tag{1}$$

- $K^{\texttt{data}}_{mt}$ is observed in data.
- $\widehat{\tilde{K}}_{mt}$ is imputed from regression.
- We recover confidence interval for parameter set using Cox & Shi (2023).

Counterfactuals: Decomposition of Growth in Data Flows

Demographics vs. Cable Network



- In 2002, improving the cable network increases total data flows by 10 times.
- A much larger effect from endowing the world with 2020 level of x_{ckt}.
- By 2013, it has a similar effect as endowing the world with the 2020 x_{ckt} (a sixfold increase).