Extracting the Costs of International Communication

by

Zachary Kiefer

University of Oregon

September 14, 2021

1 Introduction

Information frictions are a significant component of barriers to international trade. These barriers include the costs of locating buyers or suppliers, arranging the transportation and delivery of goods, and monitoring foreign market conditions, among others. An important driver of information frictions are communication costs, which are reduced by advances in communications technology allowing traders to exchange information more easily. Advances in communication technology began in the 19th century with the invention of the telegraph, and especially with the installation of trans-Atlantic telegraph cables. This advance vastly decreased the time it took to communicate between Europe and the Americas, which previously required sending information by post or courier aboard ship. Further advances, such as the telephone and fax machine, further reduced this time cost, while the decreasing cost of these technologies allowed smaller and smaller traders to acquire them.

Despite its obvious importance for understanding economic transactions, particularly on the international level, the cost of communication is difficult to measure. Communication costs are often not just about monetary costs (bills for phone, internet, webhosting, etc. services), but also about how timely, reliable and secure communication can be. These costs may also be highly correlated with factors related to infrastructure, borders, and the regulatory environment, and this makes it difficult to disentangle a robust measure of communications cost, much less its effects on international transactions. Trading firms may know that they spend money on phone and Internet bills, an email server, or the domain name for their website, but it is difficult to identify how much of those and similar costs are specifically associated with their import or export efforts. Moreover, these costs are not purely monetary: forms of communication that are slow or unreliable may impose additional costs on the firm. It is likewise difficult to extract these communications costs using macro data, as they are difficult to separate from other trade barriers without some measurement of communication.

The Internet is only the latest technological advancement in communication, and we are now seeing it expand explosively into new markets. Cellular Internet, which provides "last-mile" connections via cellular phone towers and their accompanying infrastructure, allows for Internet access to households and businesses without a landline phone connection, or even access to an electrical grid: a cellular phone can access the Internet from anywhere with a cell tower nearby, and can be charged from a gasoline generator or solar panel. In the near future, satellite Internet constellations such as the Starlink project promise to remove even the need for cellular towers. With these advances, even remote communities in the developing world have begun to gain access to the global Internet and with it, the global marketplace.

However, the Internet is unlike previous communication technologies in two important regards: firstly, it is decentralized by nature, and as such there is no central authority gathering data on the Internet as a whole. To obtain a dataset describing pricing, communication volumes, speed, etc. across the entire Internet would require piecing together data gathered by a multitude of relatively small ISPs and other networks, with the accompanying concerns about comparability and completeness. Secondly, unlike the telegraph and telephone, which were billed at rates depending on the locations of users, the Internet is typically billed at flat and uniform rates: whereas the cost of a telephone conversation between users in New York and London can be measured by a published, per-minute rate, if those same users communicate by email or video chat, the costs of that communication are obscured by the pricing structure of typical Internet contracts, which bill by the month or by the gigabyte without making distinctions based on where a user's communication will be sent.

In this paper, I overcome these obstacles to provide a new measure of internet-related communication costs. To do this, I adapt the model used in Allen and Arkolakis (2019), which was originally designed to estimate transportation costs but which models the transportation network in a way that mirrors the structure of the Internet. This framework allows me to use publicly-available data on Internet communication and routing to back out the costs of using each country-to-country link in a communications network, and from these link costs, estimate the expected costs of communicating from one country to another.

This approach presents a possible solution to a persistent problem in the trade and international economics literature; namely, the lack of good measurements of communication costs in an international context. Previous attempts to measure communication costs have largely relied on proxies such as the number of telegraphs sent (in a historical context), the number of domain names registered to a country (in a macro context), or the availability of broadband Internet (in a micro context). None of these proxies combine the desirable characteristics of (i) measuring a bilateral communication cost, (ii) being available for recent time periods, and (iii) being readily available without requiring significant effort to simply gather and prepare the data. In contrast to these proxies, I extract communication costs by first using publicly-available data to piece together a dataset measuring flows of Internet communication on a country-to-country network, after which I use the Allen-Arkolakis framework to back out communication costs which rationalize these observed flows.

The estimated costs are a bilateral measure of communication costs, and since the method used to estimate these costs relies on publicly-available data and does not require significant computing resources[1], it offers a novel and extensible way of measuring these costs, which are an important factor in many economic contexts. I demonstrate the explanatory power of the extracted communication costs in a series of gravity models of trade, and show that it has explanatory power independent of other standard measures of information and other frictions. Exploring the heterogeneity of these effects reveals that this measure has differential impacts similar to predictions of Keller and Yeaple (2013), primarily that an increase in communication costs leads multinational firms to trade more sophisticated goods, with their

[1] I have found the necessary data-processing and model-estimation scripts to run well on a cheap 2016-era consumer laptop running Ubuntu Linux, although there can be efficiency gains using more advanced hardware such as a dedicated server or parallel-processing GPU.

foreign subsidiaries. However, these communication costs have broader explanatory power; to give a handful of examples, these costs can be used to construct unilateral measures of national Internet access costs, as a component of trade costs in a Melitz model, as a proxy for monitoring costs in models similar to Head and Ries (2008) and Blonigen, Cristea, and Lee (2020).

2 Literature Review

My project connects two threads of the literature: one examining the effects that communications technology has on trade, and a second which examines the costs of trade on a network of ports or countries. However, rather than apply this transport network literature directly to trade flows, I adapt this literature to apply it to the costs of communication on a network.

2.1 Communication Costs and Trade

Previous work has addressed the effects that expanding access to Internet and other communications technologies have had on trade. Freund and Weinhold (2004) measure Internet access on a national level with a count of webhosts (e.g. "www.amazon.co.uk" or "www.disney.co.jp") registered to a country. They find that, by this measure, a 10% increase in "Internet access" by this metric stimulates trade by 0.2%. They also find that Internet access has no direct effect on the negative relationship between distance and trade values, but may have an indirect effect which in fact strengthens this relationship by increasing competition. However, the proxy used in this paper has become less applicable since 2004, because on the modern Internet, far fewer webhosts have a conveniently identifiable country-level suffix (such as ".uk" or ".jp").

Fink, Mattoo, and Neagu (2005) examine the effects of telephone communication costs on international trade, finding that these costs have a significant influence on trade patterns, which is notably more pronounced for trade in differentiated products. This is based on a cross-sectional dataset on bilateral international telephone rates from the International Telecommunications Union, which has some limitations (namely, for each "calling" country, only the rates for the 20 most popular recipient countries were available). Also, compared to 1999, a much larger share of international communication now takes place by the Internet, and data on bilateral Internet communication costs is harder to come by: unlike telephone service, where long-distance calls are commonly billed at a rate which varies by recipient country, Internet service is generally billed at a per-month rate or with a constant per-gigabyte cost. Applying Fink, Mattoo and Neagu's methodology to the Internet therefore requires some way of measuring the bilateral cost of Internet communication between countries.

Ejrnæs and Persson (2010) examine the effect of trans-Atlantic telegraph technology on violations of the law of one price, i.e. when the difference in prices for a good in two markets is not simply explained by costs of shipping. They first show that when the trans-Atlantic telegraph was completed in the latter half of the 19th century, it reduced the duration of these violations in the market for wheat, as traders in geographically separated markets (Chicago and Liverpool) learn more rapidly of the opportunity for arbitrage. They then, using an error-correction model, show that the existence of this telegraphy, allowing for pricing errors to be corrected more quickly, resulted in significant welfare gains.

Steinwender (2018) performs a similar analysis of cotton price differences between New York and Liverpool, again finding that the existence of the telegraph closed the exporter-importer price gap. Additionally, Steinwender finds the existence of the telegraph also increased export volumes, while also increasing volatility (as exporters became more responsive to fluctuations in price).

Both Ejrnaes and Persson and Steinwender are limited to considering the existence of the telegraph and not the cost of using the telegraph, presumably due to a lack of detailed data on telegraph rates from more than a century ago. They are also limited to considering a pair of locations which were centers of the trade in a commodity and were linked by telegraph at a known point in time—both of which are restrictions limiting the applicability of their approach to modern contexts.

Lew and Cater (2006) uses a gravity model to analyze the influence of the telegraph on trade among multiple countries; however, they do not use actual telegraph costs, instead using the number of telegrams sent by a country, including domestic ones. (This variable is not bilateral, and Lew and Cater's approach is to regress bilateral exports on each partner's unilateral telegrams sent.) In a sense, this measure could be a proxy for the telegraph cost within a country, but Lew and Cater interpret it as a proxy for the density (or extent) of the telegraph network within the country rather than directly addressing telegraph costs.

Allen (2014) considers the effect of cell phone access (as measured by digitizing the universe of cell tower construction permits) on inter-provincial trade in the Philippines. This paper considers a variety of measures, including price pass-through, the fraction of farmers which incurred out-of-province freight costs for their produce, and the frequency of simultaneous import and export in the same province. Unfortunately, the Philippine cell-tower data used in this analysis was costly to compile, and the NGO which compiled it appears to have closed its doors, making the extant data unavailable for further work along these lines.

Leuven, Akerman, and Mogstad (2018) use data on broadband Internet access in Norway in a gravity-based exploration of trade by and among Norwegian firms. This approach treats Internet access as a firm-level characteristic, and is limited by the scope of their broadband adoption data. This data consists of survey data from a random sample of Norwegian firms, and a measure of broadband availability (but not actual adoption) among the households of every municipality in the country, which was compiled by the Norwegian government. This combination of data sources is valuable, but rare, and does not address the cost of broadband Internet, merely its availability. Because of the relative rareness of this type of data, this approach could not easily be extended beyond Norway.

Gokan, Kichko, and Thisse (2019) develop a theoretical model which finds that less expensive transportation encourages larger numbers of "integrated" firms which carry out all of their production activity in the same location (as the cost of transportation encourages production at several locations to cut down on distance-to-market), while decreased communication costs have the opposite effect, encouraging either vertical or horizontal integration since it becomes cheaper to coordinate activities from a central headquarters. However, this paper stops short of applying the model empirically, and in the last line of the paper, suggest that communication costs could instead be linked to the opening of new airline or commuter rail routes—but a more direct measurement of communication costs could be even better.

Diverging slightly from the topic of trade, Blonigen, Cristea, and Lee (2020) finds that information frictions, specifically monitoring costs, resulting from physical and cultural distance have significant negative impacts on cross-border merger and acquisition (M&A) activity. The effect is less pronounced in the manufacturing sector, owing to the lesser importance of monitoring activity, which is an important factor in the disproportionate emphasis on manufacturing in such M&A. Costs of communication are a major factor in these monitoring costs, as modern communications technology can potentially reduce the importance of physical distance when available. Such costs, however, are difficult to measure directly.

2.2 Trade Costs on Networks

Another thread in the literature addresses the estimation of trade costs on networks. This is relevant to my work, not because I will be estimating a network trade cost directly, but because the Internet is ultimately a similar kind of network. Specifically, the Internet is a communication network structured with strong similarities to the global trade networks seen in these papers, and the costs of communication on this network can be estimated using methodologies developed to estimate the costs of trade on a global trade network.

Kikuchi (2002) provides a theoretical model predicting that countries with communications networks that are interconnected (or, by extension, interconnected to a greater degree) will have a comparative advantage in the trade of business services.

Anderson and Wincoop (2004) lay out a basic framework for the estimation of trade costs using a gravity model, or from purchasing power parity data. They also present a summary of available data on trade costs, derived from customs records, surveys of national non-tariff barriers, and other sources. However, this primarily addresses "tangible" costs and barriers to trade, leaving out intangibles such as communication costs and information frictions.

An entire sub-thread of this literature deals with transportation over a defined network, with an emphasis on enabling detailed counterfactuals: notable studies in this vein include Donaldson and Hornbeck (2016), Redding (2016), Nagy (2016), Sotelo (2015), and Ganapati, Wong, and Ziv (2020). Most relevant is Allen and Arkolakis (2014), which establishes a very general framework for modeling economic activity on surfaces with highly-adaptable topology; applying this framework, Allen and Arkolakis (2019) provides a more specific framework for estimating the costs of each link in a transportation network, which is applied to the context of inter-city trade along the US highway network. The structure of the highway network is similar to that of the Internet, and the structure of this model is convenient to adapt to cases where complete traffic data (i.e. data describing the entire universe of traffic throughout a network) is hard to come by.

2.3 Endogenous Trade Costs

I also take some inspiration from the literature on endogenous trade costs, in which the costs of trading are part of an equilibrium and are determined partly by the distribution of trade flows. The costs of communication are plausibly determined in a similar manner based on the distribution of communication flows.

Endogenous trade costs are modeled in a variety of ways in the literature: Brancaccio, Kalouptsidi, and Papageorgiou (2017) model endogenous trade costs as the result of search frictions between exporters and bulk carrier ships, with trade being direct and not upon a network. By comparison, Kleinert and Spies (2011) model separate manufacturing and transport sectors, with trade costs influenced by factors such as port efficiency (Clark, Dollar, and Micco (2004), Blonigen and Wilson (2008)), bilateral trade imbalances (Behrens and Picard (2011), Jonkeren et al. (2011)), and market structures in the transport sector (Hummels, Lugovskyy, and Skiba (2009)). These factors are in turn the results of decisions made in said transport sectors.

Allen and Arkolakis (2014) and Allen and Arkolakis (2019) derive partly from this thread of the literature as well; the latter paper specifically models trade costs as being affected by congestion, a phenomenon which occurs in communications networks much as it does in physical transportation networks.

Also of note is Duranton and Storper (2008), which uses a model of industry location with endogenous transaction costs to explain a juxtaposition between rising total trade costs and falling transport costs. This model suggests that due to increased use of complex, specialized machinery, transaction costs in the form of extensive communication between machinery manufacturer and client have offset the reduced cost of actual transport. In addition to contributing to the endogenous trade costs literature, this also motivates interest in communication costs—which are likely much lower now than in 2008, thanks to advancements in Internet infrastructure and communication technologies.

2.4 Relevant Computer Science Literature

One of the key components of my approach is geolocation of Internet end-users as well as networks. Identifying a network's geographic footprint remains a thorny problem, but Rasti et al. (2010) provides a method of doing so. This method has proven intractable to combine with other elements of my methodology, and so I abstract away from their solution.

3 Data

In order to estimate the costs of communication, it is first necessary to somehow measure the amount of communication which takes place–ideally, the amount of communication activity generated by an international trade transaction. This is a complex problem: at the micro level, there is little data measuring how many emails or phone calls a trading firm sends in the process of arranging a trade, and at the macro level, it is not feasible to separate trade-related communication from other communications. Therefore, a novel approach to measuring communication will be necessary.

In the last four decades, an increasing fraction of communication has taken place by way of the Internet, thanks to email, VoIP[2] phone systems, and videoconferencing. Due to the decentralized nature of the Internet, there is no singular authority which tracks traffic throughout the entire Internet—but there are a handful of projects which track traffic at important hubs in the Internet. Data from these projects can give some insight into aspects of how communication flows through the Internet.

3.1 Internet Routing and Routing Data

The Internet is not monolithic: rather, it is composed of many distinct computer networks that have developed protocols for cooperating and connecting with each other. An important

[2]Voice over Internet Protocol, a system which permits telephones to transmit their audio signal by Internet instead of over conventional phone lines.

part of these protocols is the approach taken to routing information over the Internet: every Internet-connected device contains instructions for how it can communicate with every other Internet-connected device. In the case of an average home computer, smartphone, or smart TV, the instructions are simple—the device just passes the user's data to another, better-informed device on their Internet Service Provider (ISP)'s network, and lets them figure it out. The ISPs, however, have to have more detailed information: they have data on routes they can use to send user data to other devices on their own network, or on other networks. At large Internet Exchange Points (IXPs), which are highly-connected datacenters where multiple ISPs and other networks connect to each other, this data is very detailed, and very complete: it contains listings of routes which can be used to communicate with the vast majority of Internet-connected devices in the world.

The Oregon Route Views Project has been collecting this routing data from a handful of major IXPs since 2003, with most of the contributors providing a snapshot of their routing data every two hours. This data is organized into observations of "blocks" of devices that can be reached by a route. Each observation provides, among other information, an ordered list of unique identifying numbers for the networks that the route passes through[3]. A brief description of the raw data is presented in Appendix A1.1.

Briefly, the routing data describes a collection of routes, the end-user devices that they connect to, and the sequence of unique networks (ISPs, IXPs, etc.) that the route passes through. These networks can be identified from a unique ID, and can furthermore be roughly geolocated, to at least a country level, based on information provided when registering and receiving that ID. Using this data, it is possible to turn the list of network IDs into a list of countries, and thus determine the sequence of country-to-country links that would be used in the route.

It is possible (and indeed quite common) that the routing data contains a multiplicity of routes: this is partly a precaution against service disruptions, e.g. a route being cut off due to a backhoe hitting a buried cable. The administrators of notable Internet entities (Internet Service Providers, Internet Exchange Points, large datacenters, etc.) are responsible for

[3]It would perhaps be better to have an ordered list of the individual devices the route passes through, but sadly, this level of detail is not available in the data.

making decisions about the routing of traffic which originates in or passes through their network, which includes deciding which of a multiplicity of routes to actually use. In practice, the Internet is so large that the administrators will not make personalized decisions on which routes to use to reach every possible destination device, but instead design a formula or process by which a computer can select the "best" route. The most widely-used criterion is route length: in the vast majority of cases, the selected route will be the most direct one: not necessarily the physically shortest route, but rather the route which passes through the fewest intermediary networks. The exceptions occur as a result of idiosyncratic variations which are not observable in this data: to give a simple example, if the administrator of network A has an old college friend at network B, they may be able to get favorable terms making routes passing through network B less costly even if they are not the most direct.

In order to make the Internet work, the distinct networks that constitute the Internet are configured to share routes with their neighbors: when a network identifies a new route which it can use to communicate with a destination device (and this route is best, or close enough to be notable), it informs neighboring networks that it has this route. The existence of this new route may then impact the neighbors' routing; if the newly-identified route gives the neighbors a new, best or near-best route, they will then announce it to their own neighbors, and the process continues. Networks with many neighbors, such as Internet Exchange Points, can select their routes from a much wider set of options, and it can generally be assumed that such a network has perfect information about the routes available to them, and has chosen the best possible route (i.e. there are no routes which an IXP would strictly prefer to use but does not know about).

In my empirical exercise, I focus on routing data from one particular collector, that being the Equinix Chicago facility.

3.2 Internet Communication and Trace Data

This routing data provides insight into how information would be routed through the Internet, but does not address the volume of such communication that takes place[4]. For

[4] To use an analogy, routing data is similar to driving directions for a road network, but it does not contain information on how heavily those directions are used.

information on communication volumes, I turn to the Center for Applied Internet Data Analysis (CAIDA)'s Anonymized[5] Internet Traces Dataset. A brief description of the raw data is presented in Appendix A1.2.

Since 2008, CAIDA has taken periodic "snapshots" of the traffic flowing through high-speed Internet backbone links. This data consists of observations of individual "packets" of information transmitted over the Internet, including the origin IP address of the packet, the destination IP address, and the size of the packet. Each snapshot captures roughly an hour of packets, and the snapshots are taken irregularly, but several times per year in the period that I'm focusing on.[6]

What makes this particular dataset useful is that one of the sources of this trace data is the Equinix Chicago facility, which is also a contributor to the ORVP. By matching this facility's trace data with its contemporary routing data, I can determine which links of the network each packet would likely use, and how much traffic they would create on those links. I can then aggregate to the link level in order to construct measurements of the amount of traffic originating from Equinix Chicago on each link in the network—which is a core component of my model. However, this is also somewhat of a limitation, because similar matching datasets are uncommon: the model I use will rely on the conjunction of routing and trace data from the same source, or hypothetically from related sources for which it can be argued that the routing data represents the true routing of the packets in the trace data. It is for this reason that I must assume that Equinix Chicago is representative of the rest of the US, rather than simply using additional data sources to get a more complete picture.

Although I am only using data from Equinix Chicago, there is nothing special about this particular source, other than the fact that it makes trace and routing data from overlapping time periods readily accessible. Similar matching datasets could be obtained in the future by partnering with similar organizations in other countries, allowing this model to be applied in a broader context.

[5]The anonymization referred to here obscures the *exact* identities of senders and recipients of data, but in a way that still allows it to be geolocated. For further information, see Appendix A1.3.

[6]To contextualize the size of this dataset, one snapshot takes up roughly 100 GB in its compressed form, and contains observations of roughly 20 billion distinct packets. The content of the packets themselves are not included in the data, merely their "metadata."

3.2.1 First-Party vs. Third-Party Communication Flows

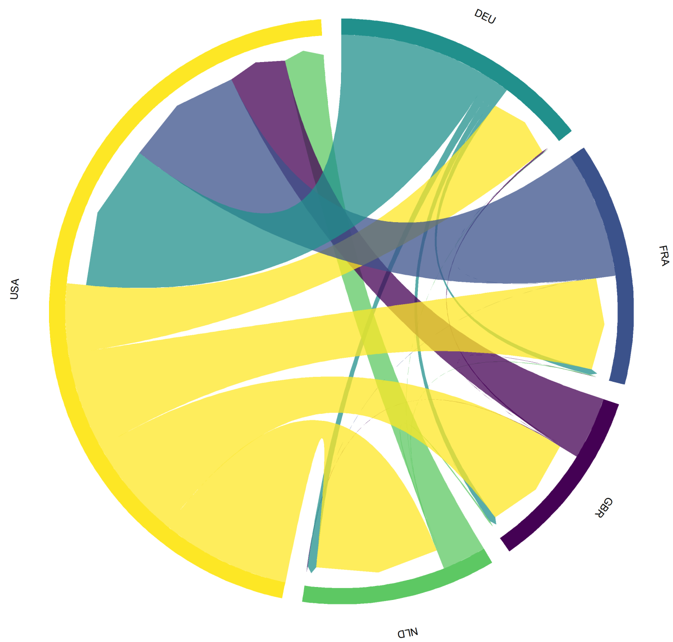
Although the Equinix Chicago facility is located on the global Internet backbone (the skeleton of long-distance, high-bandwidth communication lines that facilitate most international communication), it is still a US-based facility, and communication within the US is heavily overrepresented in the trace data. I discard observations of intra-national communication, which is unlikely to leave the US to begin with. A visualization of the flows among the five most communicative countries[7] is shown in Figure 1. Each arrow in this diagram represents an aggregate flow of international communication, with the width of the arrows indicating the size of the flow. As can be seen here, the communication data from Equinix Chicago captures an implausibly small amount of third-party communication (i.e. flows that neither originate nor end in the US): compare the size of the Germany-US (DEU-USA) flow with that of the Germany-Netherlands (DEU-NLD) flow. Considering that Germany and the Netherlands are neighbors and notable trading partners, and share a linguistic and cultural background, the amount of communication captured here is far too low. This is as expected, since Equinix Chicago can only capture communication which passes through the US-based facility, and Germany and the Netherlands have much more direct routes for their communication. It does, however, mean that this data will be best used in a model restricted to US-origin and US-destination communication, and several parts of my empirical analysis focus on this context.

3.3 Analogy to Physical Transportation

It is perhaps easiest to explain what this data captures using an analogy to the transportation of physical goods. A common question in the trade literature is how to determine the costs of trade among various locations in a transportation network: this may be country-to-country, port-to-port, or even city-to-city, depending on context. Perhaps the ideal dataset for such an application consists of two parts: measurements of trade (where goods start out, where they end up, and how valuable they are), and measurements of shipping (the value of the goods transported along various links in the network, irrespective of origin and

[7] The United States, Germany, France, the United Kingdom, and the Netherlands.

Figure 1
Bilateral Communication Flows Observed from Equinix Chicago



destination).

Given these two pieces of information, it is possible to draw conclusions about the costs of the links in the transportation network. To give the simplest possible example, if there are only two routes which connect nodes A and B in the network, and the majority of goods shipped from A to B are sent along the first route, one can reasonably conclude that the first route is the less costly to use. If similar behavior can be observed for many pairs of A and B nodes, then one can begin to draw conclusions about the factors which make routes costly by comparing the characteristics of the less-used routes.

The data in this ideal dataset could be obtained from commerce and transportation authorities, such as UN COMTRADE, the US Department of Commerce (DOC), and/or the US Department of Transportation (DOT). Unfortunately, there is no analogue to these authorities when it comes to data on Internet communication: the many distinct networks comprising the Internet may collect relevant data within their own borders, but there is no central authority which aggregates this data or ensures that the entities collecting it use a standardized process. Therefore, in this analogy, suppose that there is no federal DOT or

DOC.

Even with this restriction, it is possible to get a partial picture of where road traffic occurs by conducting a survey of drivers in a single location. Suppose that I survey drivers as they leave Eugene, Oregon, asking each one where they are driving to (analogous to the CAIDA trace data). I can then get directions to their destination from Google Maps, Apple Maps, or a variety of other sources, and record the sections of road which these directions say to drive on (analogous to the ORVP routing data). Having done this for a large number of drivers, I can then count the number of times that a driver from Eugene will (probably) use each section of road in the US highway network. While this would only produce a measurement of where traffic originating in Eugene occurs, a more complete picture could be obtained by repeating the procedure in a densely-populated or heavily-traveled area, e.g. Portland, LA or New York, analogous to IXPs.

3.4 Counting Internet Users

One straightforward way to employ this data is to simply count the number of Internet-connected devices that a collector has a working route to. This can be simply done by geolocating each block of IP addresses observed in the routing data, then summing the size of each block in a country.

This measure is not a perfect measurement of the amount of communication originating from a country, as it can be affected by variations in the number of devices per user (considerably higher in developed countries) and the intensity with which a device is used (difficult to measure, but also likely higher in developed countries). However, it has one advantage over comparable measures (such as a count of the number of IP addresses officially registered to a country), as only devices which have been connected to the Internet relatively recently, and therefore have an IP address, will be observed in the data: IP addresses which have been allocated to a country but which are not in service will not be counted. Furthermore, a count of IP addresses per country will be necessary in constructing my measure of Internet traffic.

Table 1 provides summary statistics for the number of Internet-connected devices observed in the routing data, in both total and per-capita terms. Although there are idiosyncratic variations, the general trend is for both of these measures to increase over time.

Table 1
Summary Statistics: Internet-Connected Devices by Country and Year

Variable	Levels	n	Min	$\widetilde{x}$	$\bar{x}$	Max	s
Total IPs (thousands)	2016	55	0.256	89.166	2845.529	73850.879	10824.490
	2017	55	0.512	102.439	3023.085	74454.051	11027.094
	2018	55	1.030	165.245	3181.892	76016.611	11274.412
	2019	55	4.101	224.647	3340.394	78533.905	11651.439
	all	220	0.256	141.582	3097.725	78533.905	11123.161
IPs per Capita	2016	51	0.000	0.014	0.119	1.280	0.277
	2017	51	0.000	0.022	0.129	1.283	0.284
	2018	51	0.000	0.026	0.139	1.446	0.300
	2019	51	0.001	0.030	0.209	3.879	0.617
	all	204	0.000	0.024	0.149	3.879	0.395

3.5 Procedure for Constructing Link Traffic Measurements

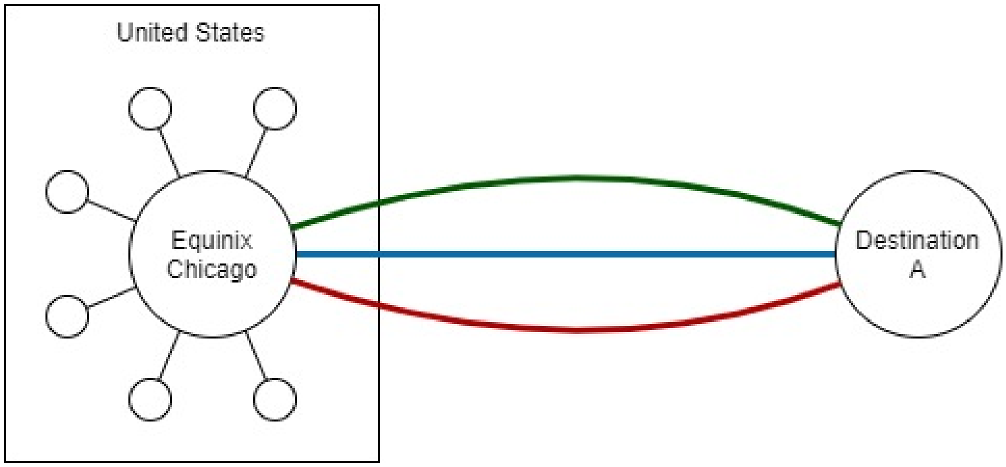
While the count of IP addresses is useful, it does not actually measure communication, or provide insight into which links in the global Internet are heavily used. To construct a measure of traffic, I couple the Equinix Chicago trace and routing datasets together, approximating the traffic generated by that facility across this network.

I begin by geolocating the origin and destination IP addresses of all observed packets, using a commercial geolocation dataset by Maxmind. I then discard all packets originating outside of the US. I denote as $IPComm_{cj}$ the total size of all the packets sent from Chicago to IP address j.[8]

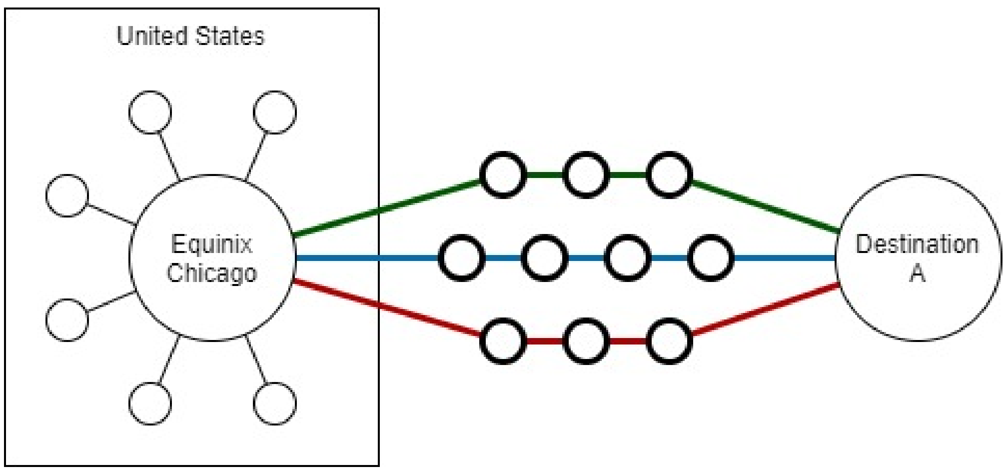
I now couple this dataset with the matching routing data: For each destination IP address j, I identify the set of routes R_{cj} which Equinix Chicago would use to reach it. Since there are frequently a multiplicity of usable routes in R_{cj}, I cut down this set using the most-direct-route criterion mentioned previously, and denote the set of most-direct routes (those with fewest intermediary networks) as R_{cj}^{min}. The selection of most-direct routes is illustrated in Figure 2.

[8]The subscript c can be interpreted as standing for Chicago, but more generally, it stands for "collector," the term used for the organizations which contributed data to the ORVP or to CAIDA. In a hypothetical application using routing and trace data from multiple collectors, c may then be used to index collectors.

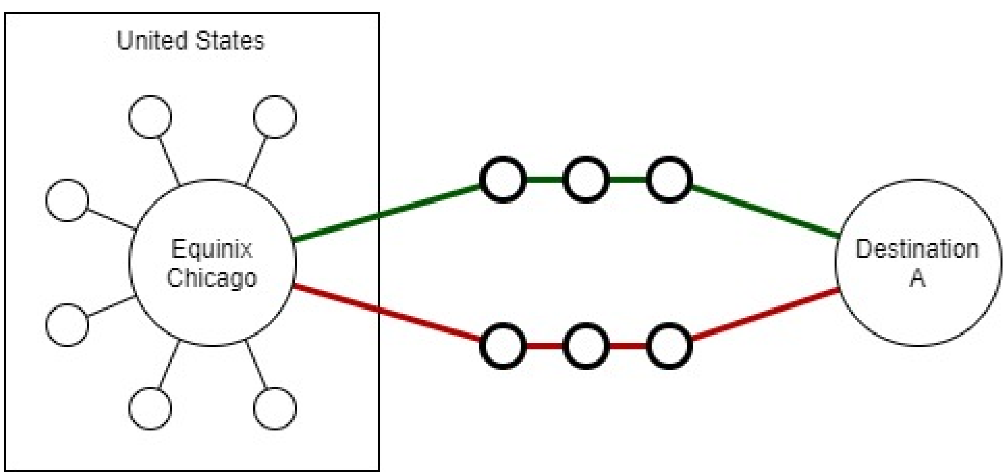
Figure 2
Selection of Most Direct Routes



(a) All routes from Equinix Chicago to an Arbitrary Destination



(b) Breakdown of Routes into Links



(c) Only Most Direct Routes

Having identified the routes which are the most direct way of reaching j, it is now necessary to divide the volume of communication sent to j among them. In cases where there is only one most-direct route, this is trivial, but in many cases there is a multiplicity of most-direct routes. As the routing data does not identify which routes are chosen, and does not contain values which can be used to condition on, I simply assign each route an equal share of communication to each most-direct route as follows:

$$RouteComm_{rcj} = \begin{cases} \frac{IPComm_{cj}}{|R_{cj}^{min}|} & \text{if } r \in R_{cj}^{min} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $|R_{cj}^{min}|$ is the size of R_{cj}^{min}, or the multiplicity of most-direct routes serving j.

Since this volume of communication will be sent over each link in the route, I next denote as $Traffic_{kl}(c, j, r)$ the amount of traffic across link kl generated by communication from Chicago to j over route r:

$$Traffic_{kl}(c, j, r) = \begin{cases} RouteComm_{rcj} & \text{if } kl \in r \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

This step is illustrated in Figure 3b.

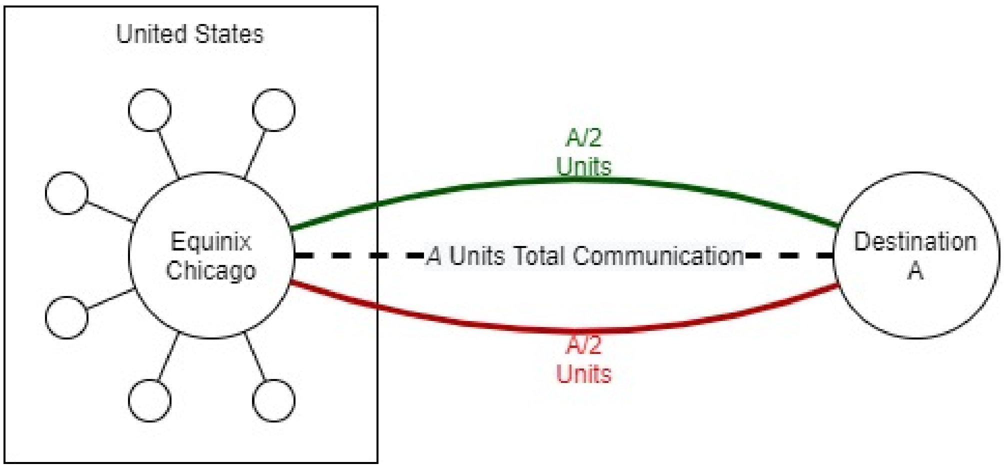
Finally, the amount of traffic originating from Chicago Equinix (or rather the US, which I take Chicago Equinix to be representative of), and present on link kl, is given by summing over destinations j and routes r:

$$TotalTraffic_{ckl} = \sum_{j} \sum_{r} Traffic_{kl}(c, j, r) \tag{3}$$

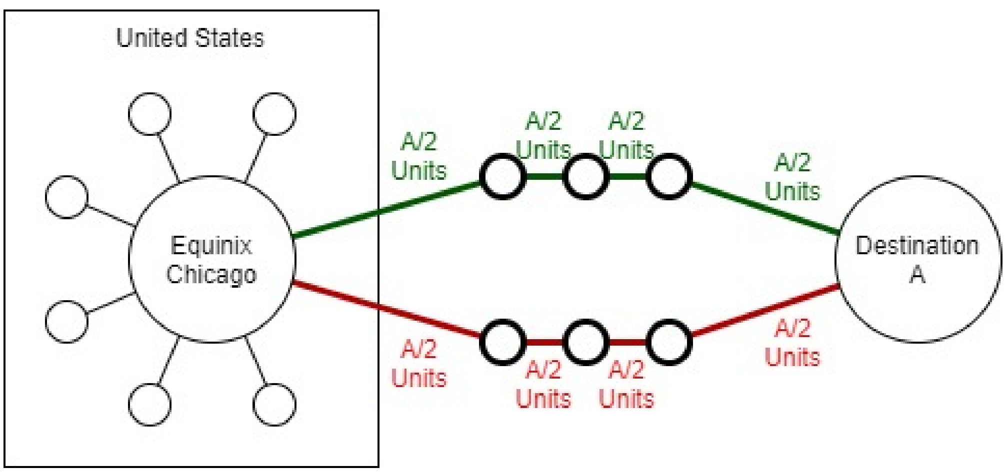
This step is illustrated in Figure 4.

This method is the most precise way of constructing a matrix of Chicago-origin traffic, but due to the enormous number of packets observed in the data, it has proven impractically computationally-intensive. Instead, in practice, I use the following, considerably faster method.

Figure 3
Assignment of Communication and Traffic

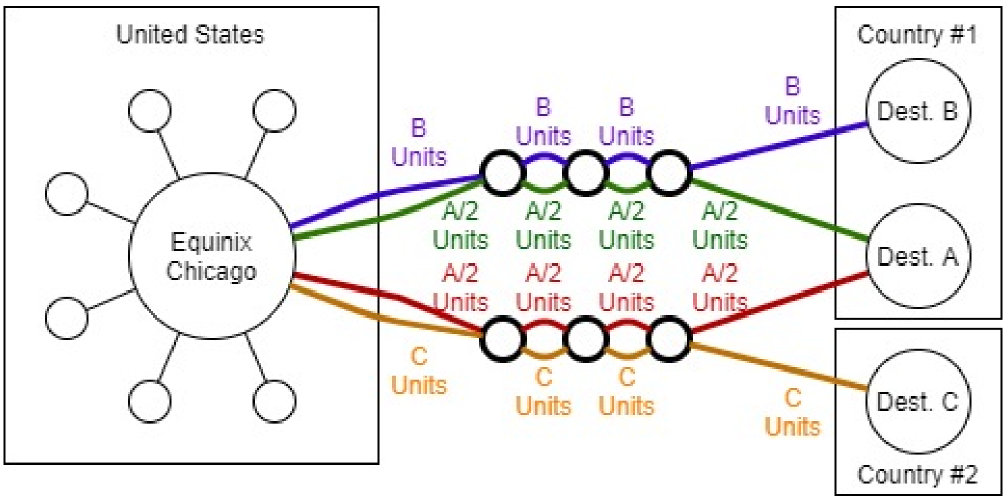


(a) Assignment of Communication Volumes to Routes

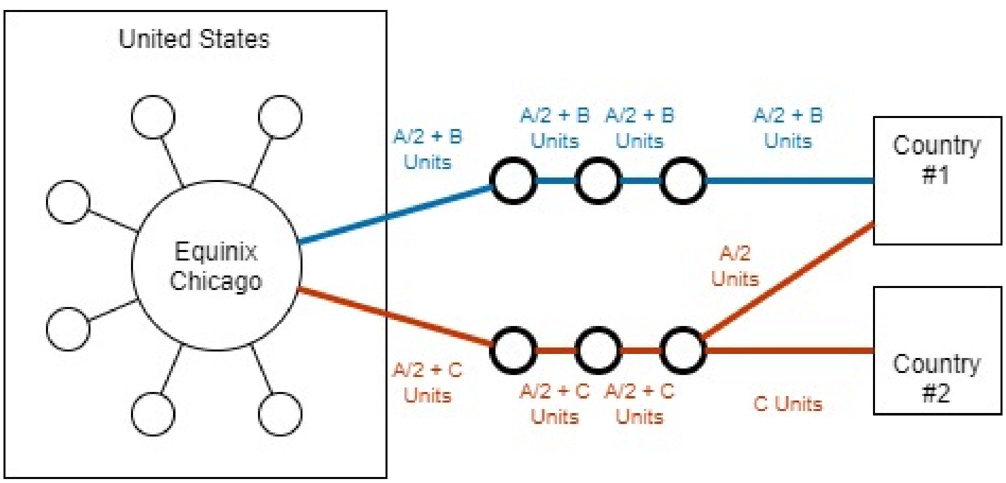


(b) Assignment of Traffic Volumes to Links

Figure 4
Aggregation of Link Traffic



(a) Assignment of Traffic Volumes to Links



(b) Summation of Link Traffic

3.5.1 Computationally Simpler Method

The computationally-intensive step in the above procedure is the matching of destination IP addresses to the blocks observed in the routing data. The routing data is not as overwhelmingly large as the trace data, but each destination IP address must be matched with an IP address block it falls into, and then with all of the routes (usually multiple) that reach that block. Given the number of IP addresses that must be matched in this procedure, it significantly reduces[9] the necessary computer time to work at a more aggregated level.

In this alternate method, I first aggregate to the destination-country level: let $CountryComm_{cj}$ be the total size of packets sent from the US to country j, with j now indexing destination countries rather than destination IP addresses. (As a side note, even if not using this simplified method, it is necessary to compute the volume of communication between all country-pairs, $CountryComm_{ij}$, as it is a key component of my model.) Also let $NumIP_j$ be the number of unique IP addresses[10] active in country j and $NumIP_\eta$ the number of addresses in block η, located in country j and observed in the routing data.

Absent any observable characteristics distinguishing IP addresses, I assume that each IP address in a country receives an equal share of communication bound for that country. I therefore assign each unique block of IP addresses in the routing data an amount of communication as follows:

$$BlockComm_{c\eta} = CountryComm_{cj} \times \frac{NumIP_\eta}{NumIP_j} \tag{4}$$

Communication and traffic can now be found similarly to the first method, but much faster owing to the per-block approach:

$$RouteComm_{rc\eta} = \frac{IPComm_{c\eta}}{|R^{min}_{c\eta}|} \tag{5}$$

$$Traffic_{kl}(c, \eta, r) = RouteComm_{rc\eta} \;\forall\; k, l \text{ s.t. } kl \in r \tag{6}$$

$$TotalTraffic_{ckl} = \sum_{\eta} \sum_{r} Traffic_{kl}(c, \eta, r) \tag{7}$$

[9]By a factor of roughly 100.
[10]Computed as described in Section 3.4.

This is, unfortunately, a simplification that loses a great deal of variation in communication flows. Given that the only way to avoid using it takes two orders of magnitude more processing time, I find it an acceptable (in fact, necessary) loss to make this approach usable for the casual user who lacks access to sophisticated computing resources.

4 Model

I adapt the framework developed in Allen and Arkolakis (2019) (from here on, referred to as the "AA framework") to estimate two sets of communication costs (the costs of using individual country-to-country links, t_{kl}, and the expected costs of end-to-end communication between countries, τ_{ij}) using this data. This framework is well-suited to this application due to the generic nature of the trade network which it models: while previous applications include road and ocean transportation networks, the structure of the Internet is sufficiently similar that it requires minimal modifications.

This model relies on two key components: a measurement of end-to-end communication between countries, and a measurement of traffic (either total traffic, or only the traffic which ultimately originates from a particular node) flowing across each link in the network. Communication can be measured by summing the total size of all packets exchanged by a pair of countries, while the traffic measurement can be constructed as described earlier.

4.1 The Nature of Costs

The τ_{ij} and t_{kl} estimated from this model are analogous to the iceberg trade costs in common usage in the trade literature. If the cost of transmitting a single unit of communication within a single network (i.e. from one device to another on the same ISP's network) is normalized to 1, then t_{kl} represents the cost of transmitting that same unit directly (without intermediary networks) from a network in country k to a network in country l.[11] The expected end-to-end communication cost τ_{ij} is likewise the expected cost of transmitting that unit all the way from a network in country i to one in country j, by whatever routes are

[11] In the special case where $k = l$, this is the cost of transmitting the unit from one network to another in the same country, which is observed to happen in the data.

optimal.

These costs do not represent costs directly paid by Internet users, but rather costs paid by ISPs, which are aggregated and passed on indirectly to users. In order to provide Internet access to their subscribers, ISPs must be able to connect subscribers to any other device on the Internet. If a subscriber wishes to communicate with another device on the same ISP's network, this is straightforward—but since ISPs are small, relative to the size of the entire Internet, it is far more common that an ISP must connect a subscriber with a device outside of their network. An ISP must therefore form some sort of agreements with other networks, to be allowed to send data outside of their own network. Such an agreement requires that the ISP pay a cost: this may be a monetary cost (an access fee to use a high-speed, long-distance cable, for example), or it may be an implicit cost: reciprocity agreements (akin to a barter transaction, in which a pair of ISPs simply agree to carry each other's communication) are common, but these come with added demands on an ISP's hardware and infrastructure, and thus indirectly impose a cost on each partner in the agreement.

Additionally, such costs need not be purely monetary: it may be more accurate to describe these costs as the cost of successfully transmitting information: if a link is unreliable, requiring repeated attempts to transmit a packet without errors, or high-latency, making it difficult to transmit time-sensitive information, this too will be captured in t_{kl} and τ_{ij}. However, these link cost measures do not map directly into a monetary cost such as "price per gigabyte" any more than iceberg trade costs map immediately into "price per 40-foot container."

The costs of constructing and maintaining an Internet link scale with distance. Longer cables naturally cost more to purchase and then install, and a longer cable also means more places that it can be hit by a backhoe, severed by a dropped anchor or bitten into by a shark[12].

τ_{ij} represents the cost to an ISP in i of transmitting information, on behalf of a user, to a recipient in j. It is uncommon for ISPs to price-discriminate on the basis of the destination of a user's communication; ISPs more commonly charge a lump-sum periodic subscription fee or a per-unit rate which does not vary depending on where information is sent. The costs

[12] A very real threat to undersea cables as recently as the 1980s.

ultimately faced by an Internet user in i could potentially be indexed by

$$C_i = \sum_j \left(\tau_{ij} \times \frac{CountryComm_{ij}}{TotalComm_i} \right) \tag{8}$$

where $TotalComm_i = \sum_j CountryComm_{ij}$. This is the expected cost of transmitting a unit of information from i, given the distribution of destinations for traffic originating in i. However, the different market conditions (competition, regulation, etc) in each country likely obfuscate these costs by inducing varying degrees of markup, which would make it difficult to draw a direct comparison between this index and data on Internet prices.

4.2 Model Environment

Let there exist a network of nodes (representing countries) connected by links (an aggregation of cables and other lines of communication). There exist a continuum of "traders", who seek to transmit information from an origin node i to a destination node j. To accomplish this, traders seek out the lowest-cost route from i to j.

A route p consists of a series of nodes $p_n,\ n = 0, 1, 2, \ldots, N$. A route from i to j begins at $p_0 = i$ and ends at $p_N = j$. The cost of such a route is the product of the costs t_{kl} associated with each link along the route,

$$\tilde{\tau}_p = \prod_{n=1}^{N} t_{k_n,l_n} \tag{9}$$

where $k_n = p_{n-1}$ and $l_n = p_n$.

However, each trader also has an idiosyncratic multiplicative cost factor $\epsilon_{p,\nu}$ of using each route, so that the cost to trader ν of using route p is

$$\tau_{p,\nu} = \tilde{\tau}_p \epsilon_{p,\nu} \tag{10}$$

Allen and Arkolakis show that, when this idiosyncratic multiplier is Frechet distributed with shape parameter θ, the traders' routing choice problem yields an analytical solution for the traffic generated by a set of link costs t_{kl}. Let A be the matrix $[t_{kl}^{-\theta}]$, and let $B = (I - A)^{-1}$, the Leontief Inverse[13] of A. Finally, let X be a matrix of observed communication flows. Then, the volume of traffic induced by these costs and communication flows is given by

$$\Xi = A \odot B'(X \oslash B)B' \tag{11}$$

where the $\odot$ and $\oslash$ operators represent Hadamard (element-wise) multiplication and division, respectively. Here, the element Ξ_{kl} is the volume of traffic flowing along link kl in the communication network.

As I am using measurements of traffic from only a single origin in the US, it is now necessary to extract from the Ξ matrix a similar measure of single-origin traffic. Allen and Arkolakis provide a convenient formula for the fraction of communication from i to j which is routed across a link kl, denoted by $\pi_{ij,kl}$:

$$\pi_{ij,kl} = (\rho \frac{\tau_{ij}}{\tau_{ik} t_{kl} \tau_{lj}})^{\theta}, \tag{12}$$

where $\rho \equiv Gamma\left(\frac{\theta-1}{\theta}\right)$. Using this formula, I am able to compute the amount of Chicago-origin traffic across links kl, given communication costs τ_{ij} and t_{kl}, and volumes X_{cj}:

$$\Xi^c_{kl} = \sum_j \left[X_{cj} (\rho \frac{\tau_{cj}}{\tau_{ck} t_{kl} \tau_{lj}})^{\theta} \right]. \tag{13}$$

Link costs t_{kl} can be parameterized as a function of observable characteristics and potentially traffic levels, if congestion is expected to affect costs. Due to the significantly different factors affecting link costs in a communications network, I impose the functional form

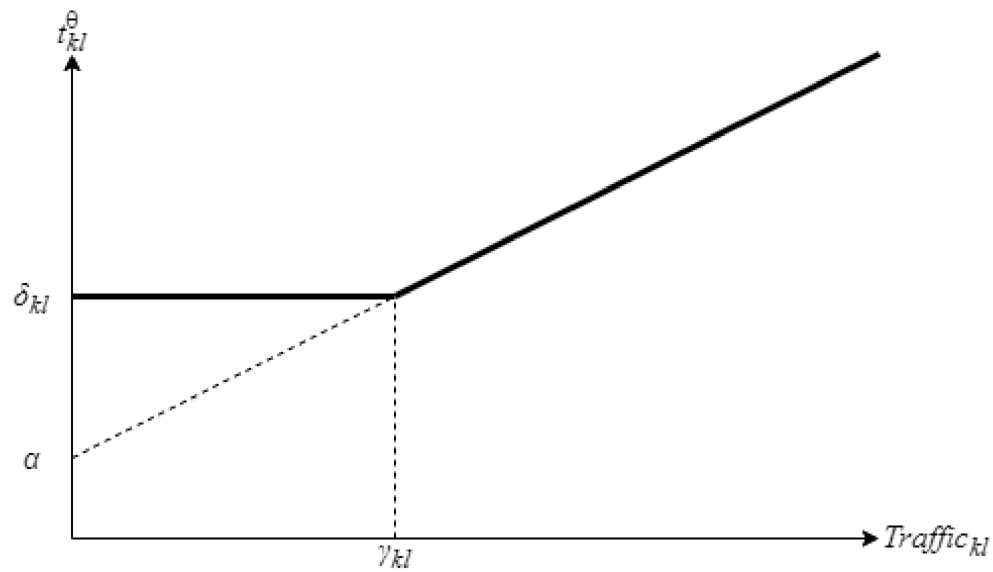
$$t^{\theta}_{kl} = \min \left[\tilde{\delta}_{kl}, \alpha + \frac{\tilde{\delta}_{kl} - \alpha}{\tilde{\gamma}_{kl}} Traffic_{kl} \right] \tag{14}$$

Here, $\tilde{\delta}_{kl} \equiv \delta Z^{cost}_{kl}$ is the ordinary cost of using link kl. This cost applies as long as

[13] In order to compute the Leontief Inverse, it is necessary for the spectral radius of A, i.e. the supremum of the absolute values of its eigenvalues, to be less than 1. In practice, this condition may be violated when the traffic matrix contains a large number of zero elements on its diagonal. The method of computing link traffic detailed above generally results in non-zero values for most of the diagonal elements of the traffic matrix, avoiding this concern.

the volume of traffic is less than $\tilde{\gamma}_{kl} \equiv \gamma Z_{kl}^{cap}$, the rated capacity of the link. Beyond this capacity, the cost of the link scales linearly: this can be seen in Figure 5. Z_{kl}^{cost} and Z_{kl}^{cap} are observables related to the cost and capacity of a link, respectively.[14] The parameters δ and γ are to be estimated.

Figure 5
Costs vs. Traffic



Given a functional form and a set of cost parameters ρ, there exists a single traffic matrix $\Xi_{pred}(\rho)$ which is rational given the costs which it induces. This traffic matrix can be found using a fixed-point algorithm which is iterated until the full traffic matrix $\Xi_{pred}(\rho)$ converges, at which point the single-origin traffic matrix $\Xi_{pred}^{c}(\rho)$ can be extracted.

The cost parameters can then be calibrated by an outer loop which searches the parameter space to minimize the distance between observed and predicted single-origin traffic,

$$|\Xi_{pred}^{c}(\rho) - \Xi_{obs}^{c}(\rho)| \tag{15}$$

Furthermore, this operation can be repeated for multiple time periods t, in order to make

[14] In an ideal world, they would be actual measurements of link cost and capacity, but no sufficiently complete dataset is readily available.

use of panel data, so that the objective function to be minimized is

$$\sum_{t} |\Xi^{c,t}_{pred}(\rho) - \Xi^{c,t}_{obs}(\rho)| \tag{16}$$

5 Estimation

I initially estimate this model using routing and trace data from Equinix Chicago in 2015-2016. Due to the well-connectedness of large IXPs like this one, I make the assumption that the routes seen from Equinix Chicago are representative of the United States as a whole. However, owing to concerns that Equinix Chicago may not accurately capture the volume of "third-party" communication flowing between pairs of countries that are not the US, I restrict this model to only US-origin and US-destination communication.[15] The available data is sufficient to work with 171 partner countries, and covers the time periods February 2015 (the earliest available period for which routing and trace data are both available), January 2016, and April 2016 (the latest available).

5.1 Link Cost Parameterization

For the observables Z^{cost}_{klt} used in the parameterization of link cost, I use data on border adjacency and the presence of undersea cables, further interacted with geographical distance (as the distance crossed by a link will naturally increase its construction and maintenance costs). The undersea cable data I obtain from a GitHub repo made available by the Telegeography Project. Z^{cost}_{kl} is thus parameterized as

$$\tilde{\delta}_{klt} = \delta_{dist}(dist_{kl}) + \delta_{adjdist}(adj_{kl} \times dist_{kl}) + \delta_{cabledist}(cable_{klt} \times dist_{kl}) + \delta_{dom}(dom_{kl}) \tag{17}$$

where $dist_{kl}$ is the centroid distance between countries k and l, adj_{kl} is an indicator taking the value 1 if k and l share a land border, $cable_{kl}$ is an indicator taking the value 1 if k and l are connected to the same undersea cable, and dom_{kl} is an indicator taking the value 1 when $k = l$ (used to set a cost for domestic links). Intuitively, the cost of an international link

[15] A method for bypassing this restriction, given appropriate covariates, is presented in Appendix A2.1

will depend in large part on the distance that the link must cover: the presence of a shared border (allowing a terrestrial cable to run directly from k to l without passing through a third country) or an undersea cable (which have generally lower maintenance costs owing to the lack of backhoes at the bottom of the ocean) merely alters the effect of distance on link cost.

Due to the scarcity of similarly detailed data on international bandwidth availability[16], I initially parameterize the bandwidth constant $\tilde{\gamma}_{kl}$ simply as a constant, i.e. $\tilde{\gamma}_{kl} = \gamma$.

I also initially allow these parameters to remain constant over time. The only cost variable which is time-varying is $cable_{klt}$, owing to a small number of undersea cables which came online during this time period.

5.2 Initial Values and Estimates

The Nelder-Mead variant used to solve the minimization problem requires a set of initial values, and unfortunately, the fixed-point algorithm in the inner loop results in an objective function with a multitude of local minima. As a result, the minimization is sensitive to the choice of initial values. In an early version of the estimation procedure, I first selected initial values by initially iterating through a discretized parameter space and selecting what were the sole set of initial values from this space which produced parameter and cost estimates satisfying two minimally-restrictive criteria:

- $\gamma_{dist} > 0$, $\gamma_{dist} + \gamma_{adjdist} > 0$, $\gamma_{dist} + \gamma_{cabledist} > 0$, and $\gamma_{dist} + \gamma_{adjdist} + \gamma_{cabledist} > 0$ so that the cost of a link between any pair of countries is increasing in distance.

- The link costs t_{kl} are all less than 10. This condition was chosen on the basis that initial runs of the model using randomly-chosen initial parameters tended to produce either costs less than 10, or extremely high values (in excess of 1000) that strained credulity in the context of iceberg costs, with little middle ground.

I have used the same initial values in successive versions of the procedure with results of similar quality in all cases.

[16] the Telegeography dataset does include some information on cable bandwidth, which is unfortunately too incomplete to rely on.

I estimate the model The coefficients estimated by the model (using only US-origin and US-destination communication) are reported in Table 3, in the Baseline column. The γ and α parameters scale with the units that $Traffic_{kl}$ is measured in (e.g. converting from bytes (B) to megabytes (MB $= 1 \times 10^6 B$) would allow the γ parameters to be scaled up and the α down by 10^6.

Table 3
Coefficient Estimates

Parameter	Baseline	Varying Gamma	Discounted Parameters
δ_{dist}	$1.189e+12$	$1.186e+12$	$1.186e+12$
$\delta_{adjdist}$	$-1.181e+12$	$-1.178e+12$	$-1.178e+12$
$\delta_{cabledist}$	$-7.279e+09$	$-7.718e+09$	$-7.718e+09$
δ_{dom}	$-1.556e+06$	$-1.556e+06$	$-1.556e+06$
$\tilde{\gamma}$	$4.274e+06$		$4.358e+06$
$\tilde{\gamma}_{Feb2016}$		$4.358e+06$	
$\tilde{\gamma}_{Jan2016}$		$4.358e+06$	
$\tilde{\gamma}_{Apr2016}$		$4.358e+06$	
α	38.84	38.84	38.84
θ	27.335	28.085	28.085
λ			0.129
κ			1.05

5.2.1 Model Fit

Overall, this model achieves a 0.159 correlation coefficient between the observed and model-predicted Chicago-origin link traffic levels. However, as seen in the scatter plot in Figure 6, the model accurately predicts volumes of traffic along a visually-distinguishable subset of links (recognizable in the plot as those which are close to the diagonal “45-degree" line), but vastly underestimates the traffic across other links. The links on which traffic is accurately predicted include both US-based and non-US-based links, and follow no discernable pattern related to cost-related variables: the geographical distance covered by these links varies widely, and there are links in this set that have shared land borders, undersea cables, both, or neither. Among US-origin links (which carry the most traffic observed from Equinix Chicago), there is a significantly greater 0.796 correlation between observed and predicted log-traffic levels.

The interpretation which emerges from these results is that this model over-costs some links in the network, resulting in the drastic underprediction of traffic on those links. Given the sparseness of the cost parameterization, I now begin to examine alternate parameterizations:

Figure 6
Observed vs. Predicted Traffic (Log Scale)

5.2.2 Estimated Costs

The distribution of link and expected communications costs estimated using this data are shown in Figure 7a. As can be seen, expected communications costs are only slightly greater than link costs, indicating that it is rare for a route to be significantly more expensive (taking into account the Frechet-distributed idiosyncratic route cost multiplier) than the direct connection with no intermediate nodes.

Additionally, it can be seen that expected communication costs for domestic links (seen at the far left of the diagrams, with costs only slightly larger than 1) are in fact less than the corresponding link costs. The interpretation of this result is that there must be some agents in this model whose draw of the idiosyncratic cost multipliers makes the cost of sending purely domestic communication out of and then back into the country less expensive than routing it purely within the country. (Or to phrase it differently, if the least-costly route for domestic communication were always the purely domestic route, the expected communication costs would be distributed with their mean at the domestic link cost.) While counterintuitive, this is actually a recognized phenomenon in Internet routing, called tromboning. It occurs when networks are not sufficiently interconnected for a direct domestic route to be cheaper than the most direct international route.

Figure 7b shows the distribution of link costs for just the “connected" links between countries with shared land borders or cable connections. As seen there, these costs fall into three rough categories: the category with lowest costs consists largely of links with both a shared border and a cable, the middle category consists mostly of links with only a shared border, and the high-cost category, which includes the right tail of the distribution, consists of those links with neither shared border or cable connection. (Links with only a cable connection are scattered throughout the middle and upper groups, but are relatively rare.) Figures 8a and 8b illustrate these breakdowns further. It should be noted that the fat right tail of the distribution, in which the costs are greater than 4, is largely composed of transoceanic links to and from the US, which are expensive due to sheer distance.

5.3 Time-Varying Bandwidth

As a robustness check, I examine whether allowing the cost parameters to vary over time impacts the results of the model. Admittedly, 2015 to 2016 is not a wide time interval, but given Moore's Law[17], it seems plausible that there could be significant reductions in communication costs year-to-year. I first allow the gamma parameter, representing the bandwidth cap of undersea cables, to vary over time, using the specification

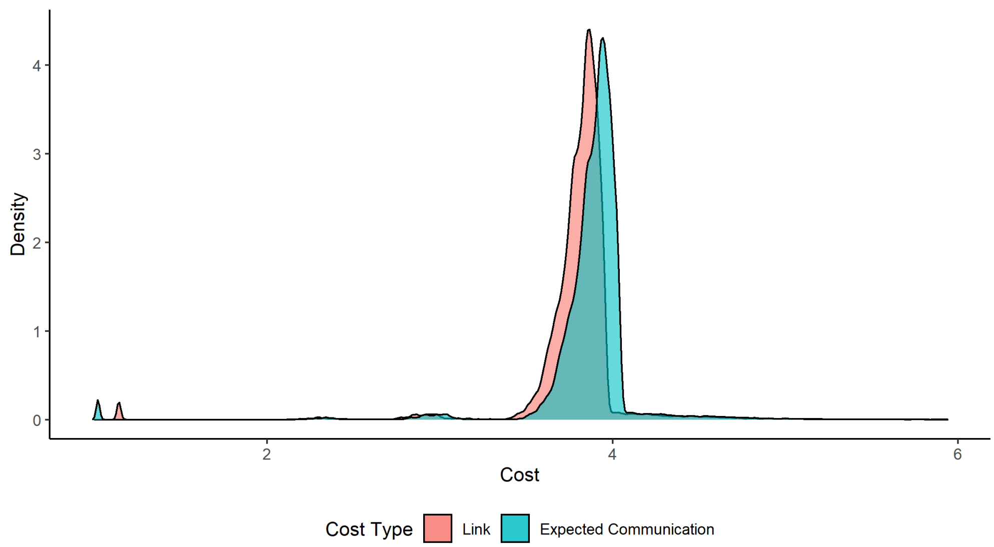
$$t^{\theta}_{klt} = \min\left[\tilde{\delta}_{kl}, \alpha + \frac{\tilde{\delta}_{kl} - \alpha}{\tilde{\gamma}_{klt}} Traffic_{klt}\right] \tag{18}$$

Parameters are reported in Table 3 under the Varying Gamma column. The $\tilde{\gamma}$ parameters are extremely similar, but not completely identical; interestingly, allowing the gamma parameters to vary over time using the same initial values has resulted in a slightly higher estimate of

[17]That computing power tends to halve in cost, or double in effectiveness holding cost constant, every 18 months

Figure 7
Distributions of Link and Expected Communication Costs

(a) Overall Distributions



(b) Connected Links Only

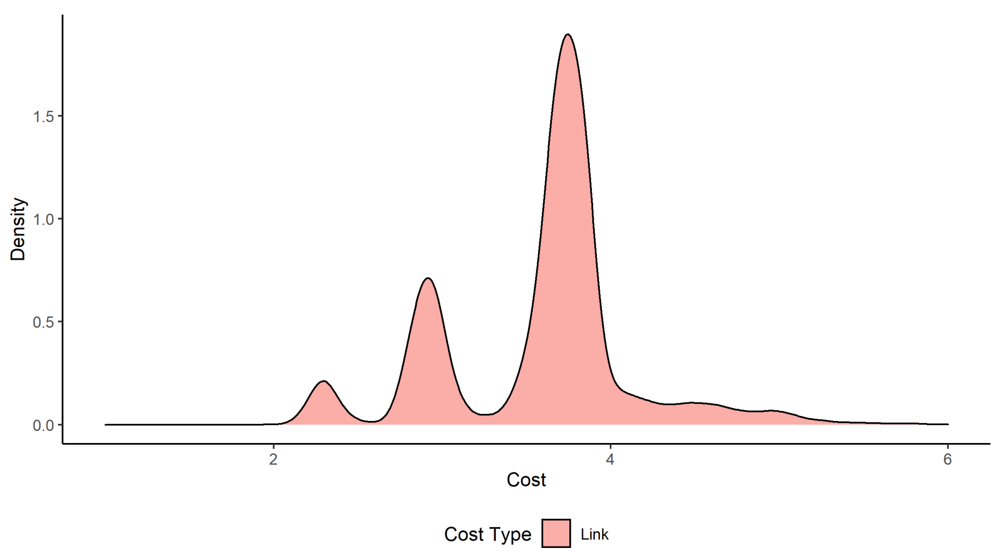
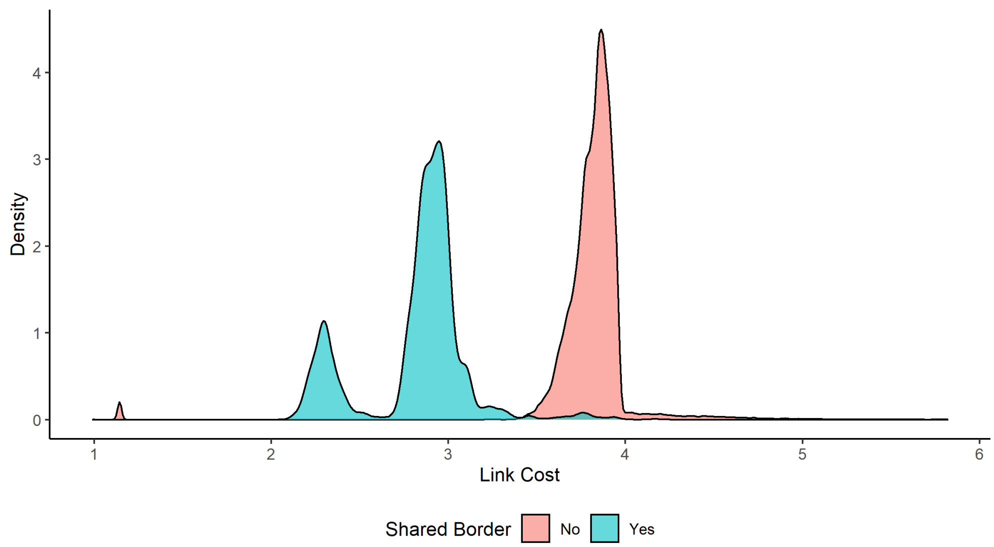
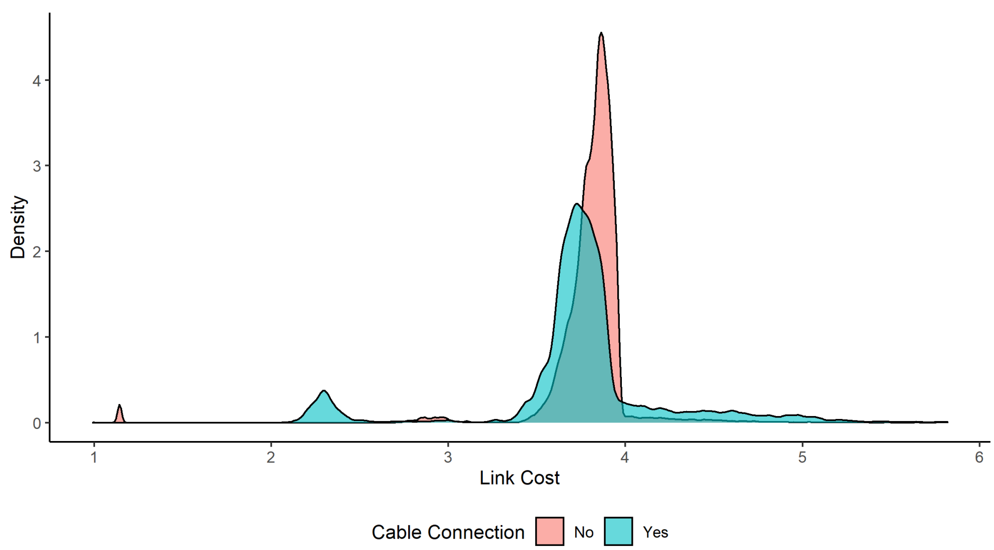


Figure 8
Detailed Breakdowns of Link and Expected Communication Costs

(a) Link Costs, Breakdown by Border Adjacency



(b) Link Costs, Breakdown by Cable Existence



θ. A comparison of estimated costs and predicted traffic is shown in Figure 9: as seen here, the change of specification has little impact on predicted traffic volumes, but slightly reduces link costs from their values in the baseline estimation. Correlation between observed and predicted traffic levels is similar to that in the baseline model.

5.4 Time as Proxy for Quality of Connection

While data on the operation cost or rated capacity of undersea cables does exist, it is not complete enough to apply in this context. However, it can be assumed that the quality of Internet infrastructure improves over time, while the cost of such infrastructure decreases. Since the data on undersea cables from the Telegeography project does include the date of activation for each cable, it is possible to use the time since the last cable on a link became active as a proxy for the quality of the link. This allows me to redefine the $\tilde{\delta}$ and $\tilde{\gamma}$ parameters, from the original parameterization, as follows:

$$\tilde{\delta}_{klt} = \delta_{dist}(dist_{kl}) + \delta_{adjdist}(adj_{kl} \times dist_{kl}) + \lambda^{t-t'}\delta_{cabledist}(cable_{klt} \times dist_{kl}) + \delta_{dom}(dom_{kl}) \quad (19)$$

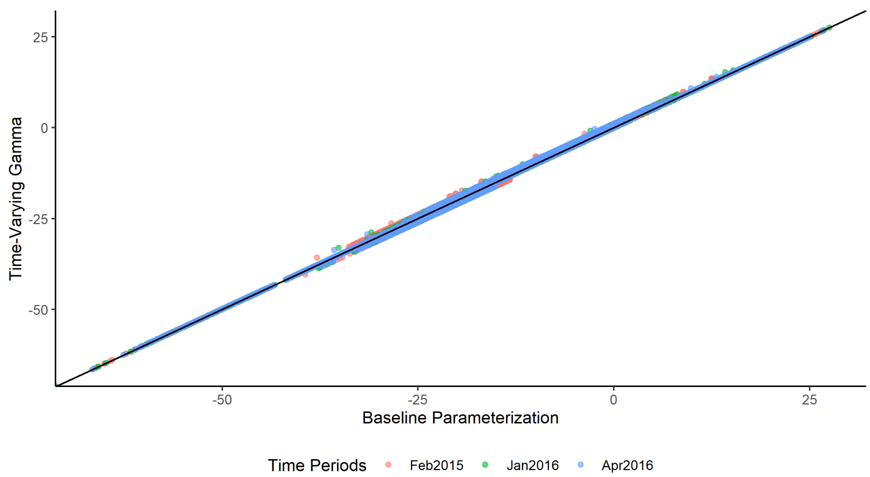
$$\tilde{\gamma}_{klt} = \begin{cases} \kappa^{t-t'}\gamma \text{ if } cable_{klt} = 1 \text{ and } adj_{kl} = 0 \\ \gamma \text{ otherwise} \end{cases} \quad (20)$$

Here, κ and λ are constants between 0 and 1, and $t - t'$ is the elapsed time, in years, between the time period t and the time at which the last undersea cable serving the link was constructed, t'. The κ and λ factors apply geometric discounting to the constants governing cost of a link equipped with an undersea cable and the rated bandwidth of such a link, respectively. This allows for operating cost to increase and rated bandwidth to decrease for links where the cables are older. This rests upon the assumption that for connections other than undersea cables, there is constant small-scale investment keeping the connection's technology up-to-date, as opposed to undersea cables which require significant lump-sum investment to build, replace, or update.

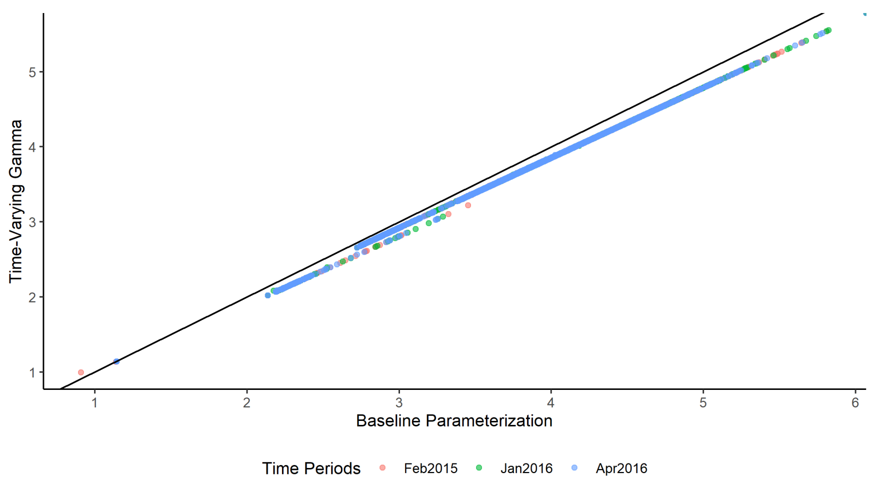
Parameters are reported in Table 3 under the Discounted Parameters column. The $\tilde{\delta}$ parameters are similar to those already estimated, but are closest to those estimated for

Figure 9
Baseline vs. Varying-Gamma Cost Specifications

(a) Comparison of Log Predicted Traffic



(b) Comparison of Link Costs



(c) Observed vs. Predicted Traffic

the time-varying Gamma model. Interestingly, the discount factor λ is quite small at 0.129, indicating that the value of an undersea cable connection drops off rapidly after coming into service—far more rapidly, indeed, than conventional wisdom such as Moore's Law[18] would suggest. The 0.129 estimate would suggest that the effectiveness of undersea cable technology to reduce communication costs doubles every 4 months, such that a 1-year old cable is roughly 1/8 as effective as a modern equivalent, which is difficult to believe, since it vastly exceeds the commonly-accepted rate of technological advancement suggested by Moore's Law.

A comparison of estimated costs and predicted traffic is shown in Figure 10: as seen here, the change of specification still has little impact on predicted traffic volumes, but due to the introduced discounting factors, vastly increases the estimated costs of the links in a nonlinear fashion. Despite this, correlation between observed and predicted traffic levels is similar to that in the baseline model.

5.5 Underprediction of Traffic

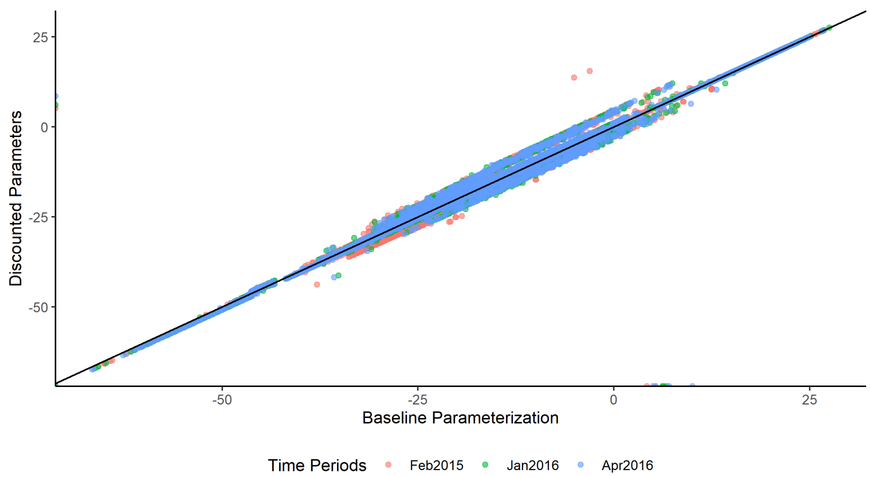
All of these cost specifications produce very similar predictions of traffic, and these predictions understate the amount of traffic on a large number of links. This, in turn, implies that the costs of these links are overestimated. Why is this the case in all three specifications? Consider the following possibilities:

1. **Omitted variables in the cost parameterization**: At its core, the functional form I use for link costs contains few variables, owing to the scarcity of complete data on the infrastructure associated with these links. There may be important cost-reducing factors which I do not have data for. In particular, the cost parameterization would be improved by a more complete dataset on undersea cable bandwidth, or even better, the bandwidth of terrestrial cables crossing land borders. Such data would allow for γ, the parameter governing link bandwidth, to be given a more nuanced functional form than the constant or discounted-constant value it takes in my specifications.

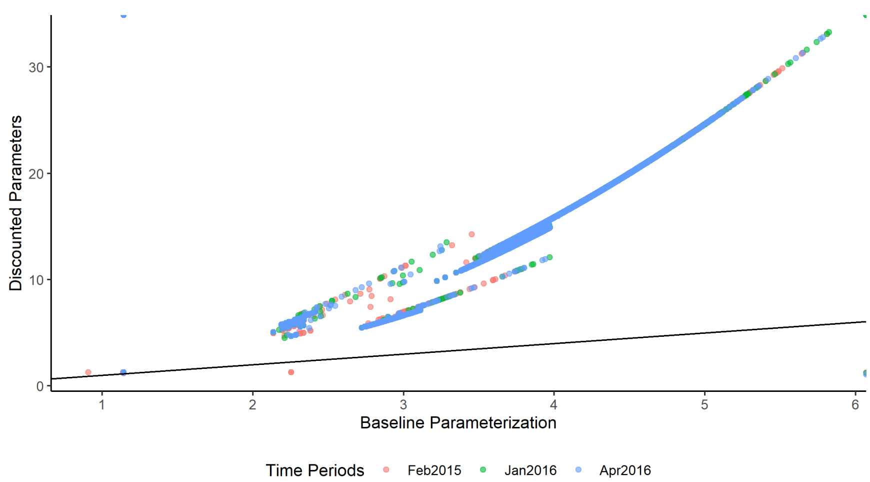
[18] A observation by noted engineer and former Intel CEO Gordon Moore that the number of transistors on a silicon chip doubles every 1-2 years, but often generalized to mean that computing power or the general effectiveness of computing technology doubles in that period.

Figure 10
Baseline vs. Discounted-Parameters Cost Specifications

(a) Comparison of Log Predicted Traffic



(b) Comparison of Link Costs



(c) Observed vs. Predicted Traffic

2. **Flawed assignment of communication to redundant routes:** Recall that when a multiplicity of routes exists, I assign an equal share of communication to each route, as visualized in Figure 3a. I do this due to a lack of observable characteristics upon which to base a more nuanced division of traffic (and, also, because it would take a prohibitive amount of time to parameterize this split and search for the ideal parameter values in my estimation process). However, it may be the case that, by assigning communication in this simplistic way, I am creating an observed traffic dataset which overstates the amount of traffic among some links, by assigning too much communication to routes which are only in the routing data as a redundant backup. With more detailed information about how a route is selected, it would be possible to refine the method by which communication is assigned to routes.

3. **Internet entities undercost some links:** A third possibility, and one that I do not place any particular emphasis on, is that the routes chosen by Internet entities are not necessarily optimal, or are optimal given some constraint which I do not model. If the routes observed in the routing data are not themselves optimal for the simplified environment I model, then the traffic I compute from it would also not be an optimal distribution of traffic. This, again, might be fixable with improved information from the providers of the routing data, as it might be possible to model the factors which affect routing choice as part of the cost function.

6 Explanatory Power Applied to Trade Volumes

With the expected communication costs estimated, I now turn to applying them in a straightforward application: a gravity model of international trade. Using trade data from COMTRADE and the communication costs estimated using non-adjusted data, I estimate the following simple gravity models:

$$log(Trade_{ijt}) = \beta_0 log(dist_{ij}) + FE_i + FE_j + FE_t + \epsilon_{ijt} \tag{21}$$

$$log(Trade_{ijt}) = \beta_1 \tau_{ijt} + FE_i + FE_j + FE_t + \epsilon_{ijt} \tag{22}$$

$$log(Trade_{ijt}) = \beta_0 log(dist_{ij}) + \beta_1 \tau_{ijt} + FE_i + FE_j + FE_t + \epsilon_{ijt} \tag{23}$$

where $Trade_{ijt}$ is COMTRADE's measure of trade volume, $dist_{ij}$ is the same centroid distance used earlier, and τ_{ijt} is the expected trade cost extracted using my model. Results are shown in the first three columns of Table 4.

As can be seen in the table, by themselves the extracted communication costs have an interpretation similar to that of distance, i.e. as a resistance term in the gravity equation, while possessing somewhat less explanatory power. When coupled together, distance absorbs much of the explanatory power of the extracted communication cost, which is to be expected considering that distance is explicitly a factor which contributes to link costs in my parameterization of the link cost function. The coefficient on communication costs is highly significant in all models where it is included, with p-values less than $2.2 \times e^{-16}$.

Thanks to the inclusion of multiple years of data (albeit condensing both 2016 observations into one year by taking the mean communication cost), I also estimate the second and third models replacing the origin, destination, and year fixed effects with origin-year and destination-year fixed effects. I omit origin-destination fixed effects, as they would absorb the distance term, which I wish to retain for comparison. As can be seen in columns (1-2) of Table 5, this has very little impact on the estimated coefficients or explanatory power of the models.

Columns (3-4) of Table 5 repeats this exercise with trade in services replacing trade in goods. Results are qualitatively similar, although note that the elasticity of trade value, with respect to either distance or communication cost, is smaller for services than for goods.

Table 4
Regression Results: Basic Gravity Models

	Dependent variable:		
	log(goodsFlow)		
	(1)	(2)	(3)
log(dist)	-2.094^{***}		-1.976^{***}
	(0.016)		(0.021)
log(meanTauCost)		-15.675^{***}	-1.972^{***}
		(0.197)	(0.230)
Observed Flows	Global	Global	Global
Fixed Effects	Orig., Dest., year	Orig., Dest., year	Orig., Dest., year
Observations	33,154	33,154	33,154
R^2	0.760	0.698	0.760
Adjusted R^2	0.757	0.695	0.758
Residual Std. Error	2.028 (df = 32849)	2.274 (df = 32849)	2.026 (df = 32848)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5
Regression Results: Advanced Gravity Models

	Dependent variable:			
	log(goodsFlow)		log(servFlow)	
	(1)	(2)	(3)	(4)
log(dist)		-1.973^{***}		-1.417^{***}
		(0.021)		(0.029)
log(meanTauCost)	-15.855^{***}	-1.995^{***}	-7.150^{***}	-0.624^{***}
	(0.199)	(0.233)	(0.221)	(0.235)
Observed Flows	Global	Global	Global	Global
Fixed Effects	Orig.-Yr., Dest.-Yr	Orig.-Yr., Dest.-Yr	Orig.-Yr., Dest.-Yr	Orig.-Yr., Dest.-Yr
Observations	33,154	33,154	7,798	7,798
R^2	0.701	0.762	0.814	0.860
Adjusted R^2	0.695	0.758	0.797	0.847
Residual Std. Error	2.273 (df = 32558)	2.026 (df = 32557)	1.335 (df = 7140)	1.160 (df = 7139)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

6.1 Heterogeneity Analysis: Breakdown by Categories of Goods and Services

I now turn my attention to the search for deeper patterns of trade, specifically, for sectors of the economy which are affected more severely by elevated communication costs.

6.1.1 Heterogeneity in Trade of Goods

I begin by breaking down trade in goods in more detail: since the sheer number of goods classifications in my COMTRADE data makes analysis on that level difficult, I instead apply the classification of goods used in Rauch (1996), which divides goods into categories of commodities, reference-priced products, and differentiated products. Using Rauch's published concordance of SITC codes to goods categories and a dataset on US exports of goods broken down by SITC code, I estimate the models

$$log(Trade_{ijtg}) = \beta_0 log(\tau_{ijt}) \times r(g) + FE_{ij} + FE_t + FE_{r(g)} + \epsilon_{ijtg} \tag{24}$$

$$log(Trade_{ijtg}) = \beta_0 log(\tau_{ijt}) \times r(g) + \beta_1 X_{jt} + FE_{ij} + FE_t + FE_{r(g)} + \epsilon_{ijtg} \tag{25}$$

where the subscript g refers to goods classified by SITC code, and ψ is a vector of indicator variables corresponding to the three Rauch classifications, each of which takes the value 1 if good g is of that classification, and 0 otherwise. In the second model, X_{jt} is a vector of destination-year controls including real GDP and population. I again condense both sets of 2016 communication costs into an average cost corresponding to the 2016 trade data, limiting the analysis to two time periods, 2015 and 2016. This specification allows for the elasticity of trade in goods to vary depending on the degree of heterogeneity a category of goods exhibits, represented by β_0 being a vector of coefficients corresponding to Rauch classifications. I omit physical distance, as it is absorbed by the destination-time fixed effect. Results are somewhat counterintuitive: as seen in Table 6, exports of commodities (about which little communication is necessary to establish the properties of the good) are reduced the most by increased communication costs; exports of reference-priced goods are reduced to a lesser degree, and in the case of differentiated goods, the effect is non-significant.

Table 6
Regression Results: Heterogeneity by Rauch Classification

	Dependent variable:	
	log(goodsFlow)	
	(1)	(2)
$log(\tau_{ijt}) \times r(Commodity)$	-1.693^{***}	-1.616^{***}
	(0.160)	(0.165)
$log(\tau_{ijt}) \times r(Ref - Priced)$	-0.564^{***}	-0.522^{***}
	(0.151)	(0.156)
$log(\tau_{ijt}) \times r(Differentiated)$	0.092	0.138
	(0.147)	(0.152)
$log(GDP_{jt})$		1.262^{***}
		(0.199)
$log(Pop_{jt})$		1.602
		(1.353)
$log(K_{jt})$		-2.472^{***}
		(0.657)
Observed Flows	US Exports	US Exports
Fixed Effects	Destination, Year, Rauch	Destination, Year, Rauch
Observations	467,616	466,162
R^2	0.259	0.258
Adjusted R^2	0.258	0.258
Residual Std. Error	2.675 (df = 467457)	2.676 (df = 466009)
Note:		*p<0.1; **p<0.05; ***p<0.01

This is the reverse of what conventional wisdom suggests should occur, in which differentiated products, which may require significant amounts of description to convey their unique product characteristics, should suffer the most from increased communication costs. A potential explanation for this pattern can be found in Keller and Yeaple (2013), which finds that multinational firms respond to greater communication costs ("disembodied knowledge transfer costs," in the terminology of Keller and Yeaple) by importing goods from their foreign affiliates that require greater knowledge to produce (an increase in the "embodied knowledge" embedded in their affiliate imports). In other words, multinational firms may respond to increased communication costs by centralizing production of more sophisticated, i.e. differentiated, products and shipping the completed good to affiliates, rather than relying on the affiliate to finish an incomplete good which may require significant information exchange.

This result is robust to the inclusion of destination-year controls, as seen from the minimal differences between the common coefficients of Models 1 and 2 in Table 6.

I therefore couple this data with a dataset measuring imports and exports to and from the US among related parties. While the dataset does represent the universe of flows observed by the US Bureau of Customs and Border Protection, documentation on the dataset does acknowledge that importers and exporters do not always report the indicator that identifies a shipment as a related-party transaction.

Using the related-party trade data allows me to estimate the regression model

$$log(Trade_{ijtg}) = \beta_0 log(\tau_{ijt}) \times r(g) + FE_{ij} + FE_t + FE_{r(g)} + \epsilon_{ijtg} \tag{26}$$

Here, $r(g)$ is a vector of indicator variables corresponding to the three Rauch classifications, each of which takes the value 1 if good g is of that classification, and 0 otherwise. FE_{ij}, FE_t, and $FE_{r(g)}$ are origin-destination, time, and Rauch classification fixed effects, respectively.

Results are reported in Table 7. As can be seen there, the τ_{ijt} cost measure has its largest effects on related-party trade in commodities, while reference-priced goods are not affected to a significant degree, and imports of differentiated goods in fact increase as communication costs rise. This result is suggestive evidence in favor of multinational corporations shifting away from trade in commodities and towards trade in differentiated goods, which tend to be more complex and therefore embody a greater concentration of knowledge. Again, these results are robust to the addition of country-time control variables, as seen in Models 3 and 4 of Table 7.

However, it is still possible to look deeper, and couple this related-party trade data with the industry knowledge intensity measures used in Bahar, Hausmann, and Hidalgo (2014) and Bahar (2019). These measures combine worker-level information to quantity the degree of "tacit knowledge" used in industries classified by SITC and NAICS codes. Using this

Table 7
Regression Results: Heterogeneity in Related-Party Trade

	Dependent variable:			
	log(goodsFlow)			
	(1)	(2)	(3)	(4)
$log(\tau_{ijt}) \times r(Commodity)$	-3.204^{**}	-15.126^{***}	-2.765^{*}	-15.176^{***}
	(1.511)	(1.923)	(1.550)	(1.964)
$log(\tau_{ijt}) \times r(Ref-Priced)$	0.217	-1.376	0.313	-1.499
	(0.949)	(1.298)	(0.972)	(1.342)
$log(\tau_{ijt}) \times r(Differentiated)$	-0.217	2.505^{**}	-0.069	2.231^{*}
	(0.851)	(1.152)	(0.871)	(1.197)
$log(GDP_{jt})$			0.341	
			(0.667)	
$log(Pop_{jt})$			-3.744	
			(5.933)	
$log(K_{jt})$			2.827	
			(2.469)	
$log(GDP_{it})$				-0.901
				(0.956)
$log(Pop_{it})$				1.042
				(9.073)
$log(K_{it})$				0.751
				(3.711)
Observed Flows	US Exports	US Imports	US Exports	US Imports
Fixed Effects	Dest., Yr., Rauch	Orig., Yr., Rauch	Dest., Yr., Rauch	Orig., Yr., Rauch
Observations	5,685	4,715	5,408	4,541
R^2	0.536	0.520	0.532	0.512
Adjusted R^2	0.522	0.502	0.517	0.495
Residual Std. Error	2.312 (df = 5512)	2.909 (df = 4544)	2.323 (df = 5246)	2.916 (df = 4380)
Note:				*p$<$0.1; **p$<$0.05; ***p$<$0.01

data, I estimate the regression

$$\begin{aligned} log(Exports_{ijtg}) = &\beta_0 log(\tau_{ijt}) \times r(g) + \beta_1 log(\tau_{ijt}) \times log(Knowledge_g) \times r(g) \\ &+ FE_{ij} + FE_t + FE_{r(g)} + \epsilon_{ijtg} \end{aligned} \tag{27}$$

Here, $Knowledge_g$ is the industry-knowledge measure for industry g, taken from the Bahar data.

Unfortunately, coupling the Bahar data with the related-trade dataset results in a highly imbalanced panel due to missing observations; there is only one commodity good observed in the related-party trade data that can be matched to the knowledge data, which forces me to drop the commodity-goods classification to avoid multicollinearity. Once again I estimate the model separately for imports and exports, with results reported in Table 8. The estimated coefficients for differentiated goods show, firstly, that all else equal, communication costs do negatively impact trade in differentiated goods, but secondly, that this effect is reduced or reversed for differentiated goods from knowledge-intensive industries. At the mean level of knowledge-intensity (weighted by the size of the export flow), the total coefficient on $log(\tau_{ijt})$ (calculated as $\beta_0 + \beta_1 \times log(\overline{Knowledge_g})$) is 0.100 for exports and 1.308 for imports, which confirms the earlier result suggesting that differentiated goods are traded more in situations with greater communication costs.

This regression also provides a more nuanced analysis of effects on reference-priced goods, which experience a net increase in trade volume from communication costs. There is some difference between effects on exports of reference-priced goods, which are driven primarily by communication costs with no significant effect from knowledge-intensity, and on import costs, where the reverse is true. However, in both cases, the total coefficient on log_{ijt} at mean levels of knowledge-intensity is much larger than the corresponding total coefficient for differentiated goods (at 7.887 for exports and 2.595 for imports). Thus, controlling for industry knowledge-intensity, it is now apparent that, at least with trade among related parties, communication costs drive a shift away from trade in commodity goods and towards more complex goods that allow for knowledge to be embodied instead of requiring difficult and expensive international communication.

6.1.2 Heterogeneity in Trade of Services

A similar analysis can be conducted with trade in services, aided by the fact that services are grouped into vastly fewer categories by EBOPS codes, and therefore bilateral service trade data is much less time-intensive to acquire through COMTRADE. Using a dataset of services trade flows reported by 133 countries, I estimate the model

$$log(Trade_{ijtg}) = \beta_0 log(dist_{ij}) + \beta_{1g} log(\tau_{ijt}) + FE_{it} + FE_{jt} + FE_g + \epsilon_{ijtg} \tag{28}$$

Table 8
Regression Results: Heterogeneity Controlling for Knowledge-Intensity

	Dependent variable:	
	log(goodsFlow)	
	(1)	(2)
$log(\tau_{ijt}) \times r(Ref - Priced)$	7.620^{***}	-0.034
	(0.856)	(1.303)
$log(\tau_{ijt}) \times r(Differentiated)$	-8.869^{***}	-4.292^{***}
	(0.750)	(1.087)
$log(\tau_{ijt}) \times r(Ref - Priced) \times Knowledge_g$	0.053	0.516^{***}
	(0.124)	(0.185)
$log(\tau_{ijt}) \times r(Differentiated) \times Knowledge_g$	1.727^{***}	1.076^{***}
	(0.047)	(0.067)
Direction	US Exports	US Imports
Fixed Effects	Dest., Year	Orig., Year
Observations	6,075	4,943
R^2	0.651	0.607
Adjusted R^2	0.640	0.593
Residual Std. Error	2.021 (df = 5902)	2.686 (df = 4772)
Note:		*p<0.1; **p<0.05; ***p<0.01

which contains a wider array of fixed effects and allows me to have significantly more observations. A summary of results is shown in Table 9, with the full results in the Appendix.

This model again produces results which go against the conventional wisdom that increased communication costs should have a purely negative effect on the value of trade in services. Instead, I observe communication costs having a mixture of positive and negative effects on trade volumes.

When I consider the types of services which experience positive or negative effects on trade volume from communication costs, a pattern emerges.

1. One broad category of goods, which I refer to as communication-delivered goods, includes those which can be exported using the Internet or other forms of communica-

tion, or which are made significantly easier to export. This type of service includes such items as health services (e.g. via telemedicine), construction abroad (which benefits from rapid exchanges of architectural plans, etc.), auditing, bookkeeping and tax consultation (all of which involve by their very nature extensive exchanges of financial data). These goods generally experience a decrease in trade value when communication costs rise.

2. A second type of service, which I refer to as communication-produced goods, includes those which can be more easily produced with easy access to communication: this type includes computing services (such as web hosting, online payment processing, etc.), research and development, advertising and market research. These services generally experience an increase in trade value when communication costs rise, as countries with generally high communication costs find it difficult to produce these goods and services for themselves and substitute towards importing. (To give one example: a country with high communication costs would find internet hosting services expensive to produce domestically, leading consumers in these countries to host their websites abroad, in countries with lower communication costs.)

3. The third category of services are those which can be thought of as enablers of or substitutes for communication. This category is effectively restricted to transportation and telecommunication services. Physical transportation can be used to transport personnel in lieu of telecommunication, or to export goods as a substitute for exporting services, while telecommunication services naturally become more expensive as communication costs rise; as such, it is expected for the value of telecommunication service exports to rise with communication costs unless there is a price effect causing volume to decrease by a large amount. The effect of increased communication costs is erratic here, with some categories (such as air passenger transportation) seeing large increases in volume with increased costs, and others (such as sea passenger transport) seeing similar decreases in volume.

Communication-delivered goods experience negative effects on trade volumes as a result of increased communication costs—exactly what the conventional wisdom suggests would occur, since these costs make it more expensive to export such goods. The other two categories experience positive effects on trade volume instead, which on close consideration seems entirely plausible:

In the case of communication-produced goods, a country which finds itself with expensive communications will also find it expensive to produce these goods. To give one straightforward example, in a country with expensive communications, web-hosting companies will face greater costs to provide a fixed level of service (as measured by latency, reliability, etc.). Alternately, these companies may choose to produce an inferior level of service. Neither of these options lends itself to producing web-hosting services domestically, and in fact this country may become a net importer of web-hosting (i.e. individuals and firms in this country may pay to host their websites and data in countries with cheaper communications).

In the case of communication-substitute goods, the rationale behind increasing trade volumes as a result of increasing communication costs is even more straightforward: faced with communication costs making long-distance communication impossible, firms may opt to send personnel between countries (to gain first-hand experience, confer with colleagues in person, or perform complicated procedures). This is akin to the outcome described in Duranton and Storper (2008), in which it becomes cost-prohibitive to export complex machinery when communication costs are high, as the amount of physical travel necessary to convey the client's specifications for a machine becomes significant. Alternately, countries which have expensive communication may choose to specialize in producing goods, not services, which also increases the quantity of transportation services necessary.

Table 9
Most Heavily-Affected Service Sectors by EBOPS Code

Top 10 categories most positively affected by communication cost	
Transportation/Air/Passenger	7.356***
Other business services/Miscellaneous/Research and development	6.443***
Other business services/Miscellaneous/Advertising, market research, and public opinion polling	5.005***
Insurance services/Auxiliary services	4.93***
Other business services/Merchanting and other trade-related services/Other trade-related services	4.128***
Communications services/Telecommunications services	3.935***
Other business services/Miscellaneous/Legal, accounting, management consulting, and public relations/Legal services	3.851***
Other business services/Miscellaneous/Other business services	3.279***
Royalties and license fees/Other royalties and license fees	3.044***
Insurance services/Reinsurance	2.753***
Top 10 categories most negatively affected by communication cost	
Personal, cultural, and recreational services/Other personal, cultural, and recreational services/Health services	−10.31***
Transportation/Other/Passenger	−8.316***
Transportation/Other/Freight	−8.173***
Transportation/Sea transport/Passenger	−7.789***
Construction services/Construction abroad	−7.087***
Other business services/Miscellaneous/Agricultural, mining, and on-site processing services/Waste treatment and depollution	−6.996***
Government services, n.i.e./Embassies and consulates	−6.698***
Personal, cultural, and recreational services/Other personal, cultural, and recreational services/Other	−6.16***
Transportation/Other/Other	−6.024***
Insurance services/Life insurance and pension funding	−3.689***

7 Conclusions

The approach I have described allows for measures of communication cost to be extracted from reasonably-accessible data on Internet routing and communication. These measures have explanatory power when used to model trade volumes, and allow for the effect of physical distance on trade volumes to be separated from the effect of communication cost (which is affected by physical distance, but incorporates other components as well).

Analysis using this data reveals trends in how communication costs affect trade, that run counter to the conventional wisdom. Coupled with data on related-party trade and industry knowledge-intensity, these trends can be explained as a result of a substitution pattern: faced with greater costs of communication, multinational firms shift away from coordinating complex global supply chains, instead performing a greater degree of transformational work in individual countries so that institutional knowledge can be "embedded" into complex goods. Given that large portions of global trade are performed by multinationals (estimates suggest values ranging from a third to a half), this substitution pattern is a noteworthy line of future inquiry.

There remain several avenues for further work on the estimation of communication costs: supplemental data regarding Internet infrastructure remains scarce, and this scarcity restricts what variables can be used to parameterize costs. This likely contributes to the major flaw of the model, which is its tendency to underpredict traffic along links far from the source of the Internet data. However, neither the scarcity of supplemental data nor the underprediction problem represent insurmountable barriers to the use of this method as a way of measuring communication costs.

A1 Data Descriptions

A1.1 Raw Routing Data

A small excerpt of relevant fields from the ORVP routing data is provided in Table A1: the excerpted observations are five distinct routes that the Equinix Chicago facility could use to communicate with a block of devices located physically near Portland.

Table A1
Excerpt from Routing Data (Equinix Chicago, January 1, 2018, 12:00 AM)

N	IP Block	Route
155044	23.206.120.0/22	53828 6939 7922 33490
155045	23.206.120.0/22	23367 6461 7922 33490
155046	23.206.120.0/22	19653 3356 7922 33490
155047	23.206.120.0/22	293 6939 7922 33490
155048	23.206.120.0/22	19016 3257 7922 33490

Taking the first row of Table A1 as an example, this observation describes a route which allows the collector, in this case the Equinix Chicago IXP, to send information to the 23.206.120.0/22 block of IP addresses. (This notation is a shorthand which refers to the block from 23.206.120.0 to 23.206.123.255, containing 1024 addresses total) This route will, after leaving the device which collected this routing data, pass through the networks with identifying numbers 52828, 6939, 7922, and 33490. These four networks are CTS Telecom, Hurricane Electric, the Comcast network backbone, and Comcast's Portland/Spokane regional network. All four are US-based.

A1.2 Raw Trace Data

A small excerpt of the relevant fields in the CAIDA trace data is provided in table A2. The three relevant fields are the origin and destination IP addresses, which can be geolocated to determine the country of origin and destination of the observed flow, and the packet size, which measures the size of the flow in bytes.

Table A2
Excerpt from Trace Data (Equinix Chicago, April 6, 2016, 1:00 PM UTC)

Origin IP Address	Destination IP Address	Packet Size
133.87.38.108	3.137.145.218	56
70.42.44.237	65.42.255.211	530
147.73.59.126	29.188.50.86	1474
161.69.48.219	161.69.45.5	1504
137.227.47.182	221.46.221.84	1504

A1.3 Trace Data Anonymization

The anonymization referred to in the name of the Anonymized Internet Traces Dataset is a prefix-preserving anonymization algorithm, which slightly perturbs the recorded origin and destination IP addresses to preserve the privacy of the users whose communication is being described. This prevents identifying the exact users who sent or received the packets recorded, but, because the algorithm is prefix-preserving, allows the users' network to be correctly identified[19]. This is sufficiently accurate for the purposes of my model, as it is unnecessary to identify anything beneath the network level.

A2 Alternate Computational Methodologies

This appendix will describe alternate methods of computing certain measures used in my model.

A2.1 Adjustments to Third-Party Communication

Because Equinix Chicago is located on the global Internet backbone, and therefore sees significant numbers of packets which neither originate from or are destined for Chicago and/or the US, it is not unreasonable to assume that this trace data accurately measures the volume of this third-party communication. However, as discussed in Section 3.2.1, it is likely that the monitoring devices at Equinix Chicago do not capture a representative amount of the communication between, e.g., Germany and the Netherlands. Adding additional

[19] To use an analogy, it is as though the addresses of respondents to a survey were obscured by altering each respondent's recorded address to a random, but still extant, address on the same street or in the same neighborhood.

communication datasets gathered from collectors in different countries would alleviate this problem by providing additional locations from which first-party flows could be measured, but using multiple datasets would also allow for third-party flows to be predicted:

Consider, as an illustrative example, a situation in which there are two communication datasets, from collectors in the US and Canada. From the perspective of the US collector, communication from Canada to France is a third-party communication flow that would not be accurately represented in the US data. But this is a first-party flow which is presumably accurately measured in the Canada data! Using flows to and from Canada, and to and from the US, it becomes possible to create a regression model which predicts the flows which are truly third-party (not to or from either Canada or the US) based on the doubly-observed communication flows. An example for this simple, two-dataset situation is as follows:

$$Comm_{ijt}^{FP} = \beta_0 Comm_{ijt}^{US} \times USTP_{ijt} + \beta_1 Comm_{ijt}^{CA} \times CATP_{ijt} + FE_i + FE_j + FE_t + \epsilon_{ijt} \quad (29)$$

Here, $Comm_{ijt}^{FP}$ is the communication observed along link ij at time t by a first-party collector, $Comm_{ijt}^{US}$ and $Comm_{ijt}^{CA}$ are the same flows as measured by the US and Canada collectors, respectively, and $USTP_{ijt}$ and $CATP_{ijt}$ are indicator variables which take the value 1 if link ij is third-party from the perspective of the US or Canada, respectively. This allows the model to be estimated using flows which are first-party to the US but third-party to Canada, and vice versa; the model can then be used to predict truly third-party flows from the imperfect measurements in one or the other of the datasets. Extensions of this approach would include the use of an averaging mechanism, so that the predictions can use the information contained in both datasets, additional covariates, and even expansion to make use of three or more datasets.

Works Cited

Allen, Treb. "Information Frictions in Trade." *Econometrica* 82.6. DOI: 10.3982/ECTA10984 (2014): 2041–2083. Print.

Allen, Treb and Costas Arkolakis. "The Welfare Effects of Transportation Infrastructure Improvements." Working Paper Series 25487. DOI: {10 . 3386 / w25487} (2019). Web. <{http://www.nber.org/papers/w25487}>.

——."TRADE AND THE TOPOGRAPHY OF THE SPATIAL ECONOMY." *QUARTERLY JOURNAL OF ECONOMICS* 129.3. ISSN: 0033-5533. DOI: {10 . 1093 / qje / qju016} (2014): 1085–1140. Print.

Anderson, James E. and Eric van Wincoop. "Trade Costs." *Journal of Economic Literature* 42.3. DOI: 10.1257/0022051042177649 (2004): 691–751. Print.

Bahar, Dany. "Measuring knowledge intensity in manufacturing industries: a new approach." *Applied Economics Letters* 26.3. DOI: 10.1080/13504851.2018.1456643. eprint: https://doi.org/10.1080/13504851.2018.1456643 (2019): 187–190. eprint: https://doi.org/10.1080/13504851.2018.1456643. eprint: https://doi.org/10.1080/13504851.2018.1456643. Web. <https://doi.org/10.1080/13504851.2018.1456643>.

Bahar, Dany, Ricardo Hausmann, and César A. Hidalgo. "Neighbors and the Evolution of the Comparative Advantage of Nations: Evidence of International Knowledge Diffusion?" *Journal of International Economics* 92.1 (2014): 111–123. Web. <http://www.sciencedirect.com/science/article/pii/S0022199613001098>.

Behrens, Kristian and Pierre M Picard. "Transportation, freight rates, and economic geography." *Journal of International Economics* 85.2 (2011): 280–291. Print.

Blonigen, Bruce A., Anca Cristea, and Donghyun Lee. "Evidence for the effect of monitoring costs on foreign direct investment." *Journal of Economic Behavior & Organization* 177. ISSN: 0167-2681. DOI: https://doi.org/10.1016/j.jebo.2020.06.008 (2020): 601 –617. Web. <http://www.sciencedirect.com/science/article/pii/S0167268120301955>.

Blonigen, Bruce A and Wesley W Wilson. "Port efficiency and trade flows." *Review of international Economics* 16.1 (2008): 21–36. Print.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. "Geography, Search Frictions and Endogenous Trade Costs." Working Paper Series 23581. DOI: 10.3386/w23581 (2017). Web. <http://www.nber.org/papers/w23581>.

Clark, Ximena, David Dollar, and Alejandro Micco. "Port efficiency, maritime transport costs, and bilateral trade." *Journal of development economics* 75.2 (2004): 417–450. Print.

Donaldson, Dave and Richard Hornbeck. " Railroads and American Economic Growth: A "Market Access" Approach *." *The Quarterly Journal of Economics* 131.2. ISSN: 0033-5533. DOI: 10.1093/qje/qjw002. eprint: https://academic.oup.com/qje/article-pdf/131/2/799/30636260/qjw002.pdf (Feb. 2016): 799–858. eprint: https://academic.oup.com/qje/article-pdf/131/2/799/30636260/qjw002.pdf. eprint: https://academic.oup.com/qje/article-pdf/131/2/799/30636260/qjw002.pdf. Web. <https://doi.org/10.1093/qje/qjw002>.

Duranton, Gilles and Michael Storper. "Rising trade costs? Agglomeration and trade with endogenous transaction costs." *Canadian Journal of Economics/Revue canadienne d'économique* 41.1. DOI: https://doi.org/10.1111/j.1365-2966.2008.00464.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2966.2008.00464.x (2008): 292–319. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2966.2008.00464.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2966.2008.00464.x. Web. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2966.2008.00464.x>.

Ejrnæs, Mette and Karl Gunnar Persson. "The gains from improved market efficiency: trade before and after the transatlantic telegraph." *European Review of Economic History* 14.3. ISSN: 1361-4916. DOI: 10.1017/S1361491610000109 (July 2010): 361–381. Print.

Fink, C, A Mattoo, and IC Neagu. "Assessing the impact of communication costs on international trade." *JOURNAL OF INTERNATIONAL ECONOMICS* 67.2. ISSN: 0022-1996. DOI: {10.1016/j.jinteco.2004.09.006} (2005): 428–445. Print.

Freund, Caroline L. and Diana Weinhold. "The effect of the Internet on international trade." *Journal of International Economics* 62.1 (2004): 171–189. Web. <https://ideas.repec.org/a/eee/inecon/v62y2004i1p171-189.html>.

Ganapati, Sharat, Woan Foong Wong, and Oren Ziv. "Entrepot: Hubs, Scale, and Trade Costs" (2020). Print.

Gokan, Toshitaka, Sergey Kichko, and Jacques-Francois Thisse. "How do trade and communication costs shape the spatial organization of firms?" *JOURNAL OF URBAN ECONOMICS* 113. ISSN: 0094-1190. DOI: `{10.1016/j.jue.2019.103191}` (2019). Print.

Head, Keith and John Ries. "FDI as an outcome of the market for corporate control: Theory and evidence." *Journal of International Economics* 74.1 (2008): 2–20. Web. <https://EconPapers.repec.org/RePEc:eee:inecon:v:74:y:2008:i:1:p:2-20>.

Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba. "The trade reducing effects of market power in international shipping." *Journal of Development Economics* 89.1 (2009): 84–97. Print.

Jonkeren, Olaf, et al. "Endogenous transport prices and trade imbalances." *Journal of Economic Geography* 11.3 (2011): 509–527. Print.

Keller, Wolfgang and Stephen Ross Yeaple. "The gravity of knowledge." *American Economic Review* 103.4 (2013): 1414–44. Print.

Kikuchi, T. "Interconnectivity of communications networks and international trade." *CANADIAN JOURNAL OF ECONOMICS-REVUE CANADIENNE D ECONOMIQUE* 36.1. ISSN: 0008-4085. DOI: `{10.1111/1540-5982.00008}` (2002): 155–167. Print.

Kleinert, Jörn and Julia Spies. "Endogenous transport costs in international trade." *Working Paper* (2011). Print.

Leuven, Edwin, Anders Akerman, and Magne Mogstad. "Information Frictions, Internet and the Relationship between Distance and Trade." 1/2018 (2018). Web. <https://ideas.repec.org/p/hhs/osloec/2018_001.html>.

Lew, Byron and Bruce Cater. "The telegraph, co-ordination of tramp shipping, and growth in world trade, 1870-1910." *EUROPEAN REVIEW OF ECONOMIC HISTORY* 10.2. ISSN: 1361-4916. DOI: `{10.1017/S1361491606001663}` (2006): 147–173. Print.

Nagy, David. "City location and economic development." 307 (2016). Web. <https://ideas.repec.org/p/red/sed016/307.html>.

Rasti, Amir, et al. "Eyeball ASes: from geography to connectivity." Jan. 2010. 192–198. Print.

Rauch, James E. "Networks versus Markets in International Trade." Working Paper Series 5617. DOI: 10.3386/w5617 (1996). Web. <http://www.nber.org/papers/w5617>.

Redding, Stephen. "Goods trade, factor mobility and welfare." *Journal of International Economics* 101.C (2016): 148–167. Web. <https://EconPapers.repec.org/RePEc:eee:inecon:v:101:y:2016:i:c:p:148-167>.

Sotelo, Sebastian. "Domestic Trade Frictions and Agriculture." 641 (2015). Web. <https://ideas.repec.org/p/mie/wpaper/641.html>.

Steinwender, Claudia. "Real Effects of Information Frictions: When the States and the Kingdom Became United." *AMERICAN ECONOMIC REVIEW* 108.3. ISSN: 0002-8282. DOI: {10.1257/aer.20150681} (2018): 657–696. Print.