THE UNIVERSITY OF CHICAGO

CONGESTION V CONTENT PROVISION IN NET NEUTRALITY:

THE CASE OF AMAZON'S TWITCH.TV

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY

JOSE FRANCISCO TUDON MALDONADO

CHICAGO, ILLINOIS

JUNE 2018

Para Mari.

Contents

List of Figures

List of Tables

Acknowledgments

Thanks to Ali Hortaçsu, Michael Dinerstein, Dennis Carlton, Brent Hickman, and Pietro Tebaldi for their guidance and support. Financial support from the NET Institute, www.netinst.org, is gratefully acknowledged. I am also grateful to the members of the IO and applied micro working groups at the University of Chicago. Thanks to the Twitch users and its community who provided me with valuable insights of the platform.

Abstract

Net neutrality encourages content provision but also creates congestion externalities from the increase in data traffic. I study the consequences of net neutrality in Twitch.tv, a popular internet platform. Twitch is non-neutral because it gives priority to the most popular content providers by compressing their data, which makes them accessible to more consumers. I estimate a two-sided-market model that considers the interactions between content provision, its consumption, and congestion. Using an exogenous technological upgrade that increased data traffic, I identify the costs of congestion for content providers and for their consumers and, using exogenous time-series variations within panels, I identify the benefits of prioritization. I use the estimated preferences and technological parameters to study the counterfactual in which net neutrality is imposed in the platform, which requires priority to be allocated anonymously. Consumer welfare drops 5%, whereas content provision does not increase, but its average quality drops. I then consider a counterfactual rent-extractive platform that charges for prioritization under the non-neutral regime. In this case, net neutrality, which prohibits priority charges, increases content provision, but consumer welfare still drops due to lower content quality and congestion externalities.

1 Introduction

Like internet bandwidth, priority is a scarce resource. During periods of congestion, network managers improve the performance of private networks by prioritizing time-sensitive data. On the internet at large, however, net neutrality not only prohibits internet service providers (ISPs) from installing pricing mechanisms to allocate priority efficiently, but also prohibits prioritization altogether.

Net neutrality is the principle that ISPs and governments should treat all data on the internet the same, not prioritizing traffic, or charging differentially by priority status, or imposing congestion charges.[1] Proponents of net neutrality argue prioritization risks being concentrated over larger firms and that such concentration will decrease the variety of content provision, which will decrease consumer welfare. Proponents of net neutrality also worry about ISPs using second-degree price discrimination as a device to extract rents from innovators and content providers. For instance, ISPs might offer content providers the option to purchase priority, and intentionally slow down delivery of data from content providers without priority. On the margin, a small fee could deter entry of small firms and start-ups. By prohibiting prioritization, net neutrality "levels the playing field," and entry becomes easier for content providers at large.

However, critics of net neutrality argue antitrust enforcement or more limited regulatory mechanisms provide a better framework for addressing competitive concerns.[2] Moreover, by differentiating between types of traffic, network managers can avoid congestion and improve utilization and quality of high-bandwidth services. Thus, even without a formal pricing system, prioritization can boost the quality of service, which will encourage content provision and investment in innovations that take advantage of improved network performance.

[1]See Lee and Wu (2009). The website of the Federal Communications Commission, `fcc.gov`, lists the rules of the Open Internet, which went live in 2015. Recently, the FCC took its first steps to roll back their Title II rules. In this paper, website addresses are abbreviated in print, but their hyperlinks are complete.

[2]See Becker, Carlton and Sider (2010).

In recent years, policymakers and the general public have escalated the debate about net neutrality. The FCC's website has received more than 22 million comments since April, 2017.[3] Europe's telecommunications regulator held a public consultation in 2016 and gathered more than 500,000 comments in six weeks—the previous record was less than 100.[4] Published on August 30, 2016, its final guidelines strongly protect net neutrality in the European Union.[5]

Net neutrality affects investment incentives. Each minute of 2017, consumers around the world spent $750,000 online, watched 4.2 million streams on Netflix, and logged in 900,000 times on Facebook.[6] The growing traffic from content providers to end users congests and strains the networks, causing service disruptions. Internet service providers claim that by collecting fees from large content providers, they can recoup investment costs and upgrade their quality of service.[7] In 2014, Reid Hastings, the chief executive officer of Netflix, expressed concerns about the bargaining power of large broadband ISPs after Netflix agreed to pay Comcast to co-locate servers inside Comcast's network (Greenstein, Peitz and Valletti, 2016*a*). Proponents of net neutrality argue ISPs will have all the bargaining power to impose high fees, which in turn will stifle content provision. And the debate heats up as more firms invest and expand their content-delivery networks around the world. Facebook and Google began to offer free or subsidized, but restricted, internet access. Now online in 53 countries, Facebook's Free Basics was banned from India in 2016 for violating net-neutrality rules, despite being used by millions of people.[8]

In theory, ISPs' and content providers' incentives to invest could both increase or decrease with net neutrality. Economides and Hermalin (2012) consider an ISP offering a fixed pipe for time-sensitive traffic, and explicitly model congestion in the ISP's network.

[3] Docket 17-108, "Restoring Internet Freedom," `fcc.gov`. Around 150,000 of those comments were received within 24 hours after HBO's John Oliver urged his viewers to send them in (`usatoday.com`). Shortly thereafter, the FCC's website was subject to a distributed denial-of-service attack (`cnbc.com`).

[4] See `berec.europa.eu`, `medium.com`.

[5] See `savetheinternet.eu`

[6] See `weforum.org`.

[7] See Greenstein, Peitz and Valletti (2016*b*).

[8] See `info.internet.org`.

They find that charges for prioritized access can serve as a price-discrimination device, which in turn can limit content provision. At the same time, however, if the ISP can price discriminate, it has more incentives to invest in its network, because it can carry more content, which the ISP charges. Thus, the authors find the overall welfare effect is ambiguous.[9]

Economides and Tåg (2012) use a membership-fee model to show a restriction on fees on the content-provider side, that is, net neutrality, may increase social welfare but only when cross-side network externalities[10] are stronger for the content provider-side. Yet consumers are worse off. However, Caves (2012) argues that gauging if their assumptions are reasonable is impossible without data.

With the exception of Nurski (2012), virtually no empirical evidence in favor or against net neutrality is available.[11] Nurski (2012) uses UK data on consumers and ISPs to estimate the demand for internet connections. Her preliminary results suggest a fast lane to a dominant content provider would increase the welfare of consumers and the profits of both the ISP offering the fast lane and the dominant content provider. I take further steps by modeling the content provider's behavior as well as the creation of congestion externalities, and by using data of a global platform.

Motivated by the theoretical literature, this paper develops and estimates a structural model that considers the aggregate effects of congestion, explicitly modeling its formation as well as its incidence on individual incentives on both sides of the market. I model the individual decisions of content providers, which depend on their individual demands, which in turn depend on congestion and content provision.

The net-neutrality debate has many aspects. I focus on the trade-off between entry and congestion. To study this trade-off, one needs data on content provision and internet

[9] See also Choi and Kim (2010).

[10] In this context, the cross-side network externalities on an internet platform are defined as the value of an extra consumer to a content provider and as the value of an extra content provider to a consumer. See Rochet and Tirole (2006).

[11] As Greenstein, Peitz and Valletti (2016*a*) argue, "There is little support for the bold and simplistic claims of the most vociferous supporters and detractors of net neutrality."

traffic. Such data are difficult, if not impossible, to collect. Besides "last-mile" ISPs, such as AT&T, "backbone" firms and content-delivery networks offer connections between last-mile providers. These intermediary connections create a tiered structure of the internet, where networks connect to other networks through a higher-tier provider. Because of these complex, fragmented relations between users, content providers, and ISPs, obtaining a comprehensive empirical picture of content supply and demand and of traffic flows on the global network is difficult.

I will therefore use Twitch.tv as a laboratory to study how net neutrality affects consumer welfare. Bought by Amazon.com in 2014 for nearly $1 billion, Twitch is an internet platform where people broadcast live video, and where people watch the broadcasts of other people. In February 2014, Twitch was the fourth-largest source of internet traffic during peak times in the United States, at 1.8% of total traffic, behind Netflix, Google, and Apple.[12] Broadcasters streamed live video from 170 countries, spoke 43 different languages, and spanned 37 different time zones. Currently, Twitch has over 100 million unique viewers and over 1.7 million unique broadcasters per month.[13]

Twitch users face trade-offs similar to those of internet users. Viewers decide which broadcasts to watch at an opportunity cost of their time. Similarly, broadcasters must decide when to stream, and their willingness to do so depends on their audience penetration. Finally, live video is sensitive to congestion, so viewers experience disutility as the result of aggregate congestion externalities caused by internet traffic jams and bottlenecks.

Twitch is not a neutral platform. It offers its most popular channels a partnership status, whose main advantage is access to a transcoder, which allows viewers to select the video quality that best suits their internet speeds. If a video is not transcoded, viewers must watch it at its original speed, which could be too high for their bandwidth. All else equal, viewers prefer higher video bitrates, but, without a transcoder, a viewer would experience video buffering and delays when she tries to watch a video with a high

[12]See *The Wall Street Journal*, `www.wsj.com`.
[13]For further details, see `www.twitch.tv/p/about`.

bitrate. The advantage of a transcoder is reflected in higher viewership for these types of broadcasts. Neutrality would require that the platform allocates transcoders without taking into account the identity of the broadcaster.

Besides transcoding information, the data, described in section 3, feature the video bitrate of broadcasts. With bitrate information, I can calculate the noise-to-signal ratio, which is a measure of the volatility of a stream. The demand for live videos is sensitive to such volatility, which manifests as buffering, stuttering, and other anomalies.

In section 4.1, I estimate a static, discrete-choice model of demand in which viewers choose their preferred channel. The noise-to-signal ratio of a channel enters demand, as a variable that brings disutility, as well as the availability of transcoding. Identification comes from exploiting three sources of exogenous variation. First, the platform upgraded the ingest capacity of its servers four separate times during the sample period. Such upgrades decreased the incidence of congestion and, by extension, the noise-to-signal ratio of uploads. Second, the platform introduced a technological upgrade that significantly increased the upload traffic and, by extension, the noise-to-signal ratio. Third, the platform allocates transcoders to non-partnered channels when excess transcoding capacity is available. Thus, identification of the transcoder effect comes from time-series variation within channels. A final challenge comes from the fact that only those channels with high viewership and low noise-to-signal ratios will be online. I use a supply-side model (introduced in the next paragraph) to construct a control function to solve this selection problem caused by endogenous decisions. I find that demand decreases with noise and increases with the availability of transcoders.

In section 4.2, I estimate a discrete-choice model of content provision in which broadcasters decide to be online or offline as a function of their individual demands and of their partnership status. To identify the effect of viewership, I construct a control function with the residuals of a first-stage regression, based on the demand model, in which the noise-to-signal ratio and the availability of a transcoder are excluded from the second-stage.

Finally, I identify the effect of partnership using time-series variation within partnered channels, exploiting the fact that their partnership comes unexpectedly. I find supply increases with the number of viewers and with the partnership status.

To close the model, in section 4.3, I estimate a technological equation in which aggregate traffic and idiosyncratic shocks explain the noise-to-signal ratio of individual channels. Here too, selection arises because channels with high noise-to-signal ratios turn off. Moreover, because noisier channels are the first to drop out when traffic increases, traffic could become endogenous. To solve the selection problem, I use the supply-side model to construct a control function. To identify the effect of aggregate traffic, I use the technological upgrade that increased broadcasting traffic. I find aggregate congestion increases the noise-to-signal ratio of individual channels.

Finally, in section 6, I use the estimates to simulate net-neutrality counterfactuals. Using the structural model, I solve a fixed point in each simulation to find the counterfactual equilibrium levels of content provision, consumption, and congestion. For a fixed number of transcoders available, instead of allocating them to the most popular channels, I impose a net-neutrality regime by allocating them randomly. I consider the whole range of transcoder availability: from none available to 100% of channels getting one, in steps of 1%. For each fixed number of transcoders available, I perform 5,000 equilibrium simulations using a random sample of 50,000 broadcasters. The simulations suggest the following. If the platform immediately switches to net neutrality, consumer surplus drops by approximately 5% with respect to the status quo.[14] To compensate consumers, the platform would need to significantly increase its investment in transcoding servers. The average quality of content provision drops. Net neutrality brings about less congestion mainly because there is less traffic. Ignoring congestion externalities would significantly overestimate the gains from net neutrality.

In a second set of counterfactuals, I consider a rent-extractive platform. In this case, the

[14]Consumer surplus measured as a function of demanded quantity.

platform charges content providers just enough to leave them indifferent between having priority and not. Net neutrality prohibits these charges as well as prioritization. In this case, content provision increases with net neutrality. However, even in this worst-case scenario with a monopolistic platform, consumer surplus decreases with net neutrality when transcoders are scarce. If, to the contrary, transcoders are not scarce, net neutrality increases consumer surplus by restoring the channels' incentives to supply.

Intuitively, the largest gains to efficiency arise when transcoders are scarce. In this case, a social planner would allocate those few transcoders to the most popular channels. Even if rent extraction discourages content provision, the gains from prioritization outweigh the losses in content provision. If quality is relatively skewed, those initial gains in efficiency will be higher. However, as transcoders become relatively common, the need for their efficient allocation decreases, and so a social planner would switch the strategy to boost content provision.

2 Net neutrality background

Since the term was introduced by Wu (2003), legal scholars have not reached a consensus on the precise definition of net neutrality.[15] I follow Greenstein, Peitz and Valletti (2016*a*): "The most basic definition of net neutrality is to prohibit payments from content providers to internet service providers. [...] Net neutrality may also be defined as prohibiting prioritization of traffic, with or without compensation."

Consider the following example. A monopolist internet service provider mediates connections between two consumers (Ana and Bob) and two content providers (Google and Facebook). The monopolist may choose prices for both sides of the market: for consumers, a price p for connecting to the internet; and for producers, termination fees t_j, for the service of "terminating" the connection between the producer and the consumer.

[15]See Greenstein, Peitz and Valletti (2016*a*).

Focus on the case in which the ISP offers the same price p for all consumers, regardless of what the consumer does on the internet. Also assume that the contracted bandwidth is the same for both content providers and that they pay the same price.

The ISP could also offer a menu of services to Google and Facebook regarding the priority of their connections. For example, the ISP could offer two lanes for traffic: a high-speed lane, priced at f_j^h, and a low-speed lane, priced at f_j^ℓ.

A "weak" version of net neutrality would require the monopolist ISP to set $t_j = t$ for all producers j. Analogously, it would also require that $f_j^h = f^h$ and $f_j^\ell = f^\ell$. That is, weak net neutrality requires that the ISP offers the same options to every producer.

A "strong" version would require the monopolist to set $t = 0$ always, or, analogously, to set $f^h = f^\ell = 0$, essentially prohibiting prioritization. Most of the literature refers to net neutrality in the "strong" sense: termination fees and prioritization fees to be outlawed altogether.

Continuing with the example, consumers are better off if both Google and Facebook enter the ISP's network. Likewise, both Google and Facebook are better off if both consumers enter the network. That is, cross-side network externalities exist. However, the ISP's incentives may not be aligned with either side of the market. We cannot discard that the ISP may prefer to charge high fees to the producers to the point that one of them drops out. In such a case, net neutrality has the potential to be welfare enhancing by increasing the number of mutually beneficial exchanges.

Finally, a more common type of externality is in play: congestion. As more consumers and producers enter the platform, the network capacity gets congested through the increase in data traffic. For a given infrastructure, more traffic slows everyone's connection, rendering poor-quality user experience.

These issues are the trade-offs of the policymaker. Net neutrality may facilitate entry on the producer side, which increases entry on the consumer side, and a feedback loop ensues, fueled by cross-side network effects. On the other hand, entry increases congestion, which decreases demand, which decreases supply, which decreases congestion, and another feedback loop ensues.

This paper focuses on a “fast lane versus the dirt road” type of prioritization. That is, as opposed to price discrimination, the platform might provide preferential treatment to some content providers, giving their data a “fast lane” to users, so that other traffic would receive a “slow lane” and arrive later. In this case, a net-neutrality regime would impose a single speed and quality for all data, or at least would require priority to be allocated anonymously. The following section describes general characteristics of the internet and introduces the empirical context this paper analyzes.

3 Data

Collecting relevant data on the internet’s content provision and its consumption poses significant challenges. First, consider size. According to Cisco (2015), the annual global internet traffic will reach 1.1 zettabytes in 2016, or roughly 152 million years of high-definition video.[16] Thus, one needs to be parsimonious about what data to collect.

Second, consider structure. Besides “last-mile” internet service providers, such as AT&T, “backbone” firms and content-delivery networks offer connections between last-mile providers. These intermediary connections create a tiered structure of the internet, where networks connect to other networks through a higher-tier provider. Indeed, backbone internet firms are called tier 1 providers. The backbone is largely competitive. However, the tiered structure of the internet is flattening as firms vertically integrate content provision and distribution. Indeed, video is becoming increasingly pervasive on the web, currently accounting for 70% of all internet traffic, and could reach 82% by 2020 (Cisco, 2015). The shift to video has motivated content providers to expand their content-delivery networks, which will carry 64.5% of total internet traffic by 2020 (Cisco,

[16] 1 zettabyte = 10^{21} bytes.

2015). Pioneered by Google and YouTube, the expanse brings content physically closer to consumers, bypassing the higher-tiered networks.[17] This fragmented structure and the complexity of its interconnections inhibit the centralization of internet data.

Bought by Amazon.com in 2014 for nearly $1 billion, Twitch.tv is an internet platform where (1) people broadcast live video, and where (2) people watch the broadcasts of other people. In February 2014, Twitch was the fourth-largest source of internet traffic during peak times in the United States, at 1.8% of total traffic, behind Netflix, Google, and Apple.[18] Currently, Twitch has over 100 million unique viewers per month and over 1.7 million unique broadcasters per month. On average, each viewer watches 106 minutes a day.[19] Thus, Twitch is a feasible compromise that parallels the most relevant features and trade-offs internet consumers and producers face, namely, that live video is sensitive to congestion, consumers value content provision and its quality of service, and content producers value its audience penetration.

Twitch specializes in broadcasting video-game sessions and content, but other categories include entertainment, social, news, sports, animals, creativity, and poker. In 2014, Twitch already dominated the gaming-video industry, broadly defined, with 43% of the market share, and maintained an extensive presence: broadcasters streamed from 170 countries, spoke 43 different languages, and spanned 37 different time zones (see Table 2, below). In 2017, Twitch remains the largely uncontested leader in live-streaming content of any kind[20] and, as *The Economist* reports, "e-sports, in which computer gamers compete before thousands of fans in person and millions more online, is on the rise," with the number of people watching e-sports reaching 258 million in 2017.[21]

Twitch broadcasters manage one channel, registered to them when they join the platform. The platform does not charge for signing up or requesting a channel. A channel

[17] See `qz.com/742474`.
[18] See The Wall Street Journal, `www.wsj.com`.
[19] For further details, see `www.twitch.tv/p/about`.
[20] According to a recent industry report by SuperData Research, `superdataresearch.com`.
[21] Source: `economist.com`.

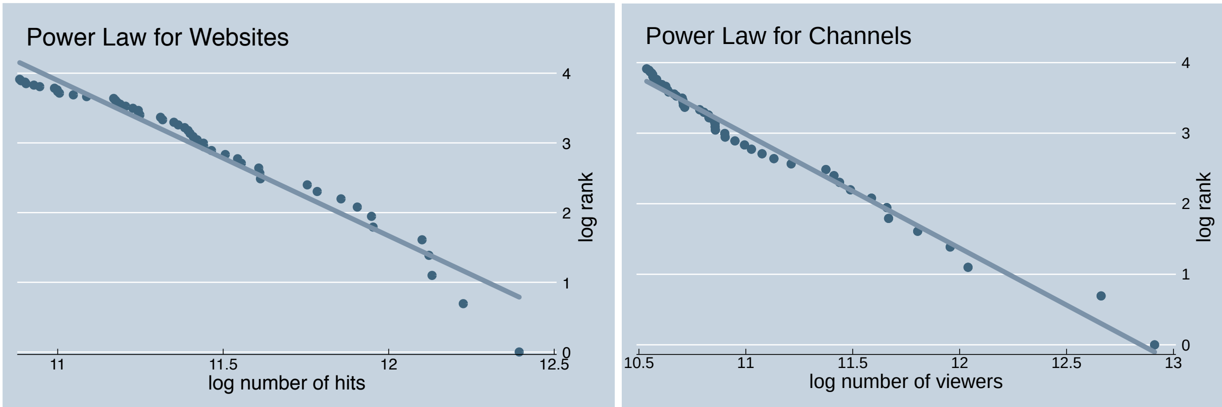
can either be online at a given time, which means that the uploader is broadcasting a live video, or offline. No videos, live or otherwise, are available when the channel is offline. Besides deciding what to broadcast, the user also decides how long to broadcast and at what average speed, measured in bits per second. When a channel is online, it corresponds to a session. The number of viewers changes during the session, and the broadcaster observes the changes in real time. Watching a channel is also free, except for the time spent watching ads, which amounts to the main source of revenue for the platform.[22]

The platform gives preference to popular broadcasters with a "partnership program." To join the program, the broadcaster needs to meet certain technical requirements, to be popular enough, and to be willing to join. The benefits of the partnership include: receiving a share of ad revenue, the ability to accept donations from viewers in the form of a subscription, and access to video transcoders, which allow viewers to select the video quality that best suits their internet speeds.[23] That is, a partnership does not carry extra costs to the broadcaster. On the contrary, partners receive a compensation from the platform. At any given moment, less than 10% of online channels have access to transcoders, which is the main advantage for popular channels, in contrast to YouTube, where all (non-live) videos are transcoded by default. In this sense, the platform is not neutral, because it gives preference to a minority of channels. This paper explores the counterfactual that imposes neutrality on Twitch by changing the distribution of partnerships and transcoders.

Using Twitch as an analogy for the internet, the most popular channels correspond to the big players such as Google or Facebook, whereas the least popular channels correspond to small websites and amateur producers. Notably, the platform exhibits a power law in the popularity of channels akin to the power law found on the internet (Newman, 2005;

[22]The data were collected in 2014. The platform has changed in the meantime, so this description applies to January-April of 2014.

[23]Transcoders were guaranteed only for partners at least until July 2015. Currently, the platform has expanded transcoding to a limited number of non-partners. Source: `blog.twitch.tv`.



Note: Power law comparison for websites (left) and broadcasters (right). Size taken as unique visits for websites and peak viewers for broadcasters. Website data from comScore ranking of June 2017 (`comscore.com`). Platform data from Pires and Simon (2015) and the author.

Figure 1: Power laws on the internet and the platform

Gabaix, 2009; Pires and Simon, 2015). Figure 1 compares the internet with the platform when popularity is plotted against ranking.

As most of the literature does, this paper will abstract from the tiered structure of the internet, so we can focus on the trade-offs consumers and producers face. In particular, this paper will be most useful when considering a monopolist internet service provider and single-homing consumers of video. Indeed, the "Federal Communications Commission found that a limited percentage of US households had access to a provider of broadband at 25 Mbps or more, and 20 percent had no access" (Greenstein, Peitz and Valletti, 2016*a*), whereas global mobile data traffic was just 5% of total IP traffic in 2015 (Cisco, 2015). Thus, single-homing and a monopolist platform seem to be close to reality. Moreover, a monopolistic platform is the best-case scenario for net neutrality to be able to protect content providers.

The raw data were mainly collected by Pires and Simon (2015) in the context of computer-communication networks, which I complemented using the Twitch REST API.[24] The data consist of the global state of Twitch channels roughly every ten minutes from

[24]The raw data can be found in `dash.ipv6.enstb.fr/dataset/live-sessions/`. The do files that I used to parse it, complement it, and analyze it are available upon request. REST API stands for REpresentational State Transfer Application Programming Interface. It is a communication method between the programmer and the platform and allows requests of data.

January 6, 2014, to April 6th, 2014. Each panel follows a broadcaster and contains channel ids, session ids, time stamps, real-time number of viewers, video bitrates, resolution, duration, country of origin, language, time zone, and some other information. More than 1 million unique broadcasters are active in about 25Gbs of data. Tables 1 and 2 present the most relevant information.

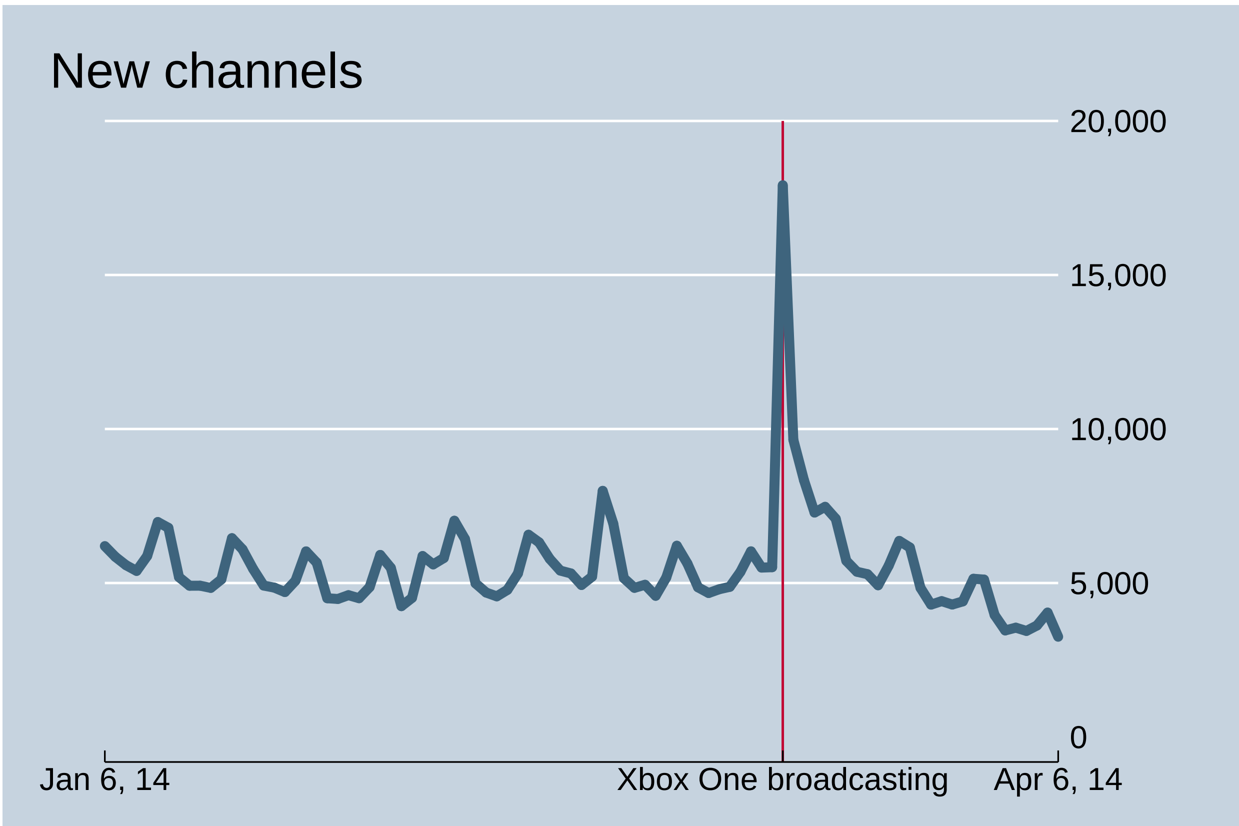
Table 1 shows the main statistics of the streaming sessions. The average broadcasting session has an average of 15 viewers, lasts for 65 minutes, and streams at a video bitrate of 1.5Mbps.[25] Gaming channels account for around 95% of the sessions, but non-gaming sessions last longer: 57 minutes versus 202. However, non-gaming sessions have lower viewership and also have less quality, as represented by a lower mean bitrate and higher noise-to-signal ratio (details follow). The main statistics do not differ much by weekend versus weekday, but a disproportionately high number of sessions occur on weekends. Finally, partnership plays a role as expected: partners have much longer sessions, better viewership numbers, and better video quality. The difference between a partnered broadcast and a non-partnered one is roughly the difference between high-definition and standard-definition.

Table 2 presents the most frequent countries of origin, languages, and time zones in the data. As the table shows, the United States and Europe are the largest markets, followed by Taiwan. Notably, the platform is blocked in China. The use of virtual private networks might mask countries of origin. However, VPNs have tell-tale signs, such as switching countries from session to session, and suffer from the disadvantage of slowing down the connection, which is essential for streaming.

Twitch users can broadcast their sessions from their personal computers or video-game consoles (e.g., PS4, Xbox One), and people can watch from any device. However, before March 11, 2014, viewers could watch broadcasts from the Xbox One, but users could not broadcast from the Xbox One. On March 11, 2014, Twitch launched the complete app for

[25] A YouTube 480p video requires around 1.5Mbps, whereas a 720p video requires around 2.5Mbps. Source: `https://support.google.com/youtube`.

the Xbox One, which allowed broadcasting.[26] This introduction of a new technology is the source of an exogenous variation that affected the broadcasters, specializing in video games, but not the viewers. Figure 2 shows the time series of new accounts for the list of channels that are found in the data. The time series peaks on the date when the Xbox One broadcasting became available, suggesting the validity of this introduction as an instrument.



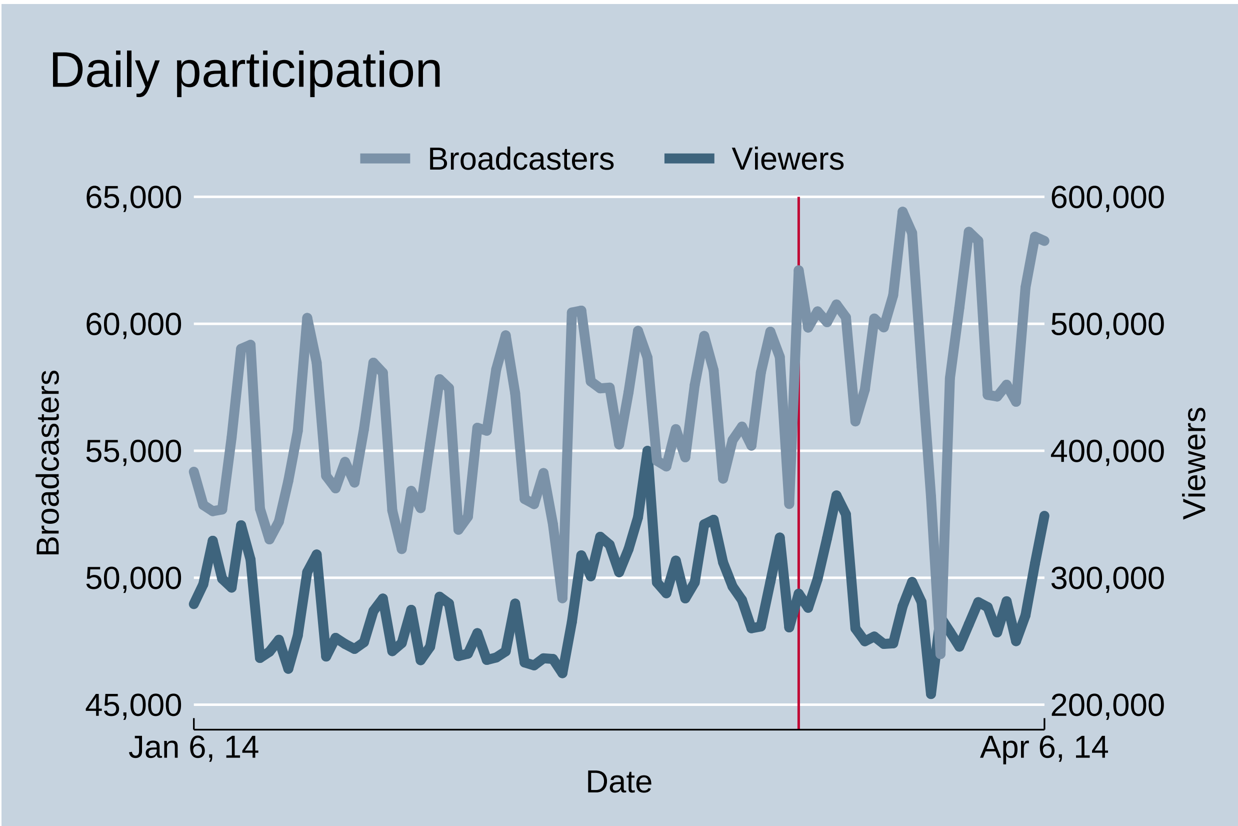
Note: Daily number of new broadcasting accounts created. Data collected by the author from the Twitch REST API . Vertical line marks March 11, 2014.

FIGURE 2: Broadcaster account creation

According to the platform's website, on March 11, 30% of broadcasters were streaming from the Xbox One, and within the first week, a total of 108,000 unique broadcasters were streaming from the Xbox One, accounting for 22% of Twitch's broadcaster base.[27] Figure 3 shows the time series of daily broadcasting and viewership. The peaks and

[26]Source: `blog.twitch.tv`, published February 25, 2014.
[27]Source: `blog.twitch.tv`, published March 31, 2014.

valleys correspond to a seasonality of one week, with peaks on weekends. On March 11, a Tuesday, broadcasting numbers jumped, confirming the press release from the platform. On the other hand, viewership lags behind and takes time to catch up. The data suggest congestion played a role.



Note: Daily number of unique channels in the data and daily maximums of concurrent viewers in the data. Peaks correspond to weekends. The vertical line is on Tuesday, March 11, 2014.

FIGURE 3: Daily broadcasting and viewership

The video bitrate is one of two key measures of the quality of a stream. All else equal, a higher average bitrate implies a better quality, because more information can be encoded in any given period of time. However, a viewer requires a high internet speed to be able to watch a live video being broadcasted at a high bitrate. Therefore, the broadcaster faces a trade-off between "viewability" and quality. The Twitch guidelines and its community suggest a target of 1.5kbps to optimize viewership, which the data confirm. On the other hand, a broadcaster whose channel features a transcoder does not face this trade-off, because all her viewers can dial down the source speed to match an appropriate bitrate to their connections. Thus, a transcoded channel will choose the maximum speed its connection can handle. For these reasons, I treat the choice of video bitrate as exogenous.

On the other hand, a high video bitrate does not guarantee a better quality for viewers and could in fact decrease the quality of a video. To decode a stream, a computer expects consistency from the source, specially if the encoding protocol is Constant BitRate, which Twitch requires. Network congestion at any level creates variation in the streamed data (due to lost data packets and bottlenecks), and is exacerbated by the size of the data packets being sent. This volatility, which is the second key measure of quality, manifests into video buffering or stuttering.[28] Evidence suggests delays and loss of packets, such as those created by volatility, significantly decrease the end-users' quality of experience while watching a video stream (Pankert, Faggiano and Taga, 2014).

This paper measures volatility with the noise-to-signal ratio of the stream's video bitrate. The ratio is a form of coefficient of variation and is composed of standardized absolute deviations from the mean. All else equal, a low noise-to-signal ratio implies better quality and is largely beyond the control of the broadcasters.[29]

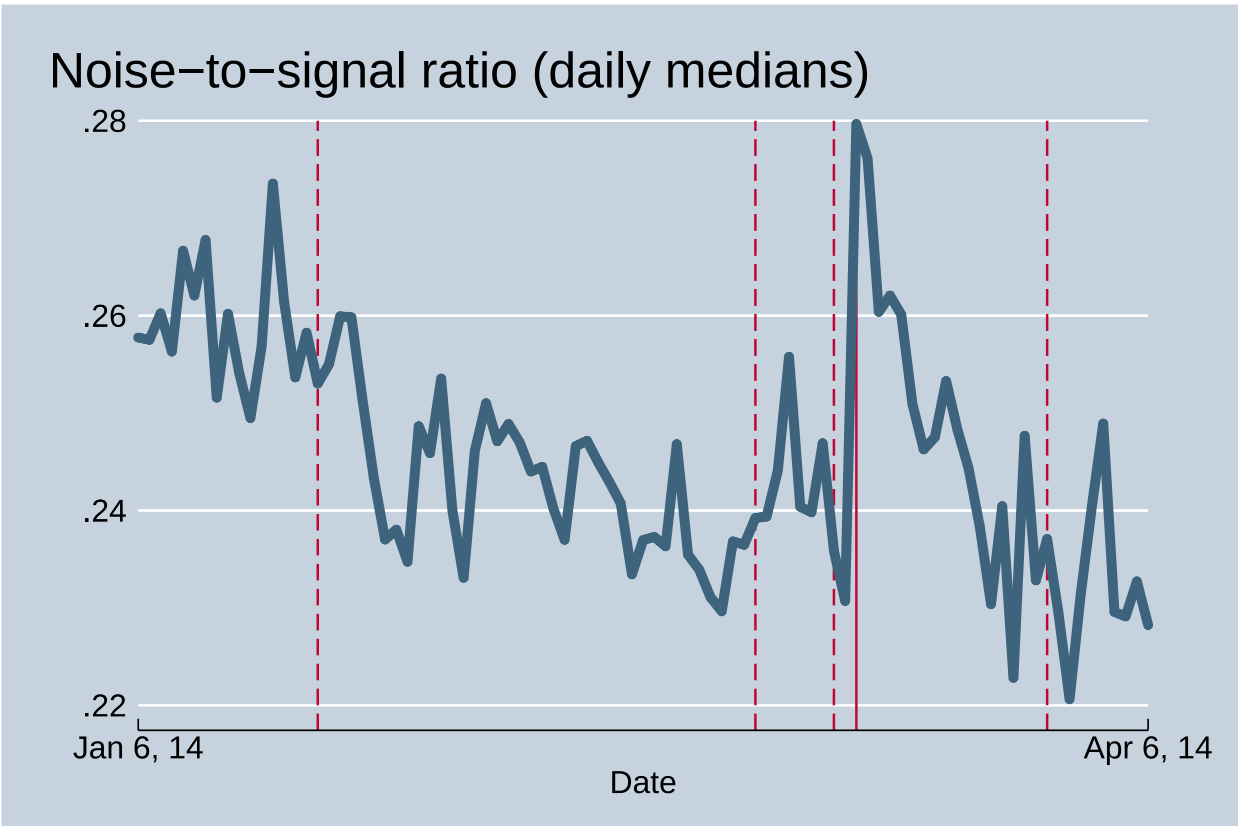
Notably, the platform expanded its infrastructure four times during the sample period by upgrading their points of presence (PoP).[30] The expansion of PoPs increases the ingest capacity of the network and improves user experience in general. Indeed, the data show decreases in noise following each installment, after controlling for confounders. Figure 4 depicts daily medians of noise-to-signal ratios and shows a remarkable increase in noise when Xbox One broadcasting was enabled.

Because the platform's physical network has a fixed throughput capacity, one would expect the noise-to-signal ratio to increase as more viewers and broadcasters participate

[28] See also `help.twitch.tv`.

[29] For applications of the noise-to-signal ratio in the engineering sciences, see Huynh-Thu and Ghanbari (2008) or Shivaldova, Winkelbauer and Mecklenbrauker (2014). Huynh-Thu and Ghanbari (2008) discuss the scope of the noise-to-signal ratio to assess video quality.

[30] Source: `blog.twitch.tv`, published March 28th, 2014.



Note: Daily median of noise-to-signal ratios. Dashed lines signal the platform's point-of-presence upgrades. The solid line signals the Xbox One broadcasting shock.

Figure 4: Quality of service and point-of-presence upgrades

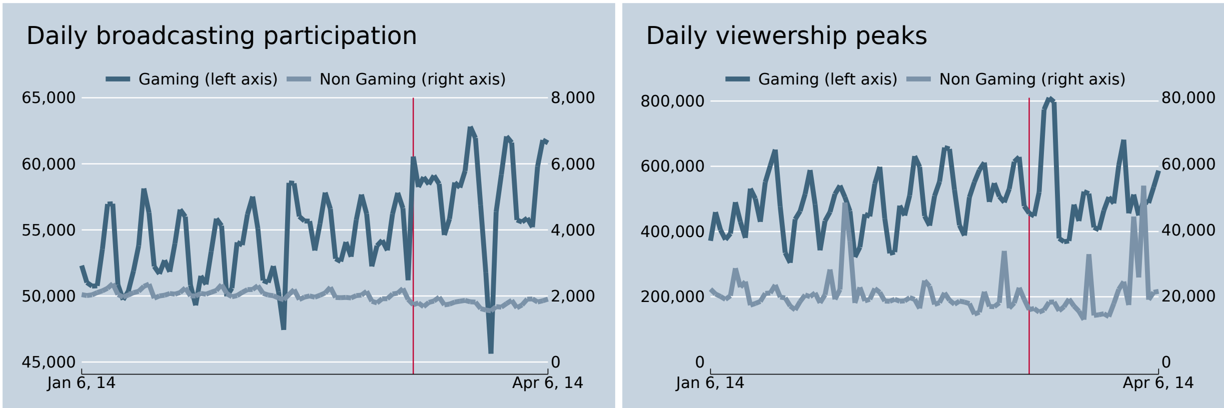
in the platform. Indeed, increasing the uploads and downloads worsens queuing delays, packet loss, and connection blocking. However, its effect is confounded by selection, because viewers and broadcasters choose not to participate in the platform when it is congested, because they value the quality of the broadcast. The noise ratio is thus an endogenous variable: less noise attracts more people, and more people creates more noise. The identification strategy is composed of a combination of instruments to disentangle the effect of noise over participation and vice versa. Section 4 discusses the empirical strategy.

Table 3 compares the platform's state before and after the Xbox One acquired broadcasting capacity on March 11, 2014. The data were aggregated in 10-minute windows. The table shows concurrent sessions, concurrent viewers, a congestion measure, and the duration of the sessions. First, the overall number of online channels increased from 5,266 to 5,562. By category, the number of online gaming channels increased but the number of online non-gaming channels dropped. The Xbox One broadcasting introduction only benefited gaming channels, but by itself does not explain the drop in non-gaming channels. Only when congestion is taken into account can the drop be explained.

Table 4 shows the results of a difference-in-differences estimation, taking the non-gaming category as a control group. Figure 5 shows a graphical representation. The estimates confirm that the introduction of the new broadcasting technology increased the number of channels by 372 and the number of viewers by 7,486 for the segment of the market in the gaming category. Controlling for time trends or weekends has virtually no effect on the estimates. However, as explained above, negative externalities from congestion over the control group could lead us to overestimate the effect. These issues are addressed below with the structural model.

3.1 Data limitations

The platform gives preference to channels in three main forms: promoting the channels to partners, giving them salience through features, and providing them with transcoders.



Note: Broadcasting: daily number of unique broadcasters. Viewership: daily maximums of concurrent viewership. Vertical line marks March 11, 2014.

Figure 5: Comparison of Gaming and Non-gaming time series

First, in the data, I do not observe if a channel is a partner, but I observe whether the channel has enabled a subscription button, which is only available for partners. Viewers can purchase a subscription for $4.99 per month to support a partner, in which case they will not have to watch ads on that channel.[31] Because this benefit is a minor one for the viewer, subscriptions are best described as tips. Twitch does not release the split on subscription revenues with partners, but it can range from 80/20 to 40/60 depending on their popularity.[32] From interviews with the Twitch community, I find partners virtually always enable the subscription button. Indeed, the purpose of partnership is to get access to subscriptions.[33] Thus, the data identify partners with near perfect accuracy.[34] Second, the data fully identify if a channel is featured on the homepage.

Third, the number of transcodes is not publicly available. A transcoding server is a scarce resource because it uses specialized hardware to encode a single stream into

[31]Other benefits of a subscription include: access to specially designed emoji, special icons that display alongside a subscriber's name, special alerts, and exclusive chatrooms.

[32]Twitch/partner. Source: `businessinsider.com`.

[33]If a partner does not have subscriptions, she could still receive a share of the ad revenue generated by her channel and individual donations. However, partners prefer subscriptions to donations, because a subscription could lock a viewer into a monthly commitment.

[34]The exact number of partners is not public, but the press reports that on March 11, 2014, there were "over 5,100 partner channels on Twitch, up from 3,386 at the end of 2012," (`fastcompany.com`); see also `polygon.com`. By July 2014, the number rose to 6,500 partners (`pastemagazine.com`), by December 2014, there were 9,000 (`washingtonpost.com`), and by October 2015, about 12,000 (`engadget.com`).

several in real time. Reportedly, a transcoding server that costs about $1,000 can provide transcoding to four live channels at once.[35] A popular channel, broadcasting HD video to 250 viewers, would cost $2,000 per hour using a third-party cloud-based transcoding service at current market rates.[36] Only partnered and featured channels are guaranteed a transcoder. However, the platform uses excess capacity to provide transcoders to the most popular non-partnered live channels.[37] In the data, the total number of concurrent partnered and featured channels peaks at 334, meaning Twitch at least has capacity to provide transcodes to 334 channels. I assume this number is a good approximation of the true (soft) upper bound of transcodes for three reasons: first, Twitch reserves a few transcoders as a buffer for an unexpected surge in partnered channels (which peaks at 248 concurrent partners), so the platform tries to be conservative; second, the observed maximum of 334 occurred only once and during peak hours; third, if the required number of transcodes suddenly increases, ingest servers can supplement transcoding, albeit inefficiently. Results are not sensitive to this assumption.

Finally, when Twitch allocated transcoders to non-partner channels, the broadcasting session had to be interrupted.[38] Thus, a tell-tale sign of receiving a transcoder is an interruption, which is observed in the data. With this information, I can reconstruct to a high degree of accuracy the channels that had a transcoder even though they were not partnered or featured. Because having a transcoder is essentially random from the point of view of the channel, the identification strategy can be summarized as assuming that having a transcoder is exogenous conditional on a fixed effect. The following section presents the model and the identification argument in detail.

[35] `reddit.com/r/Twitch`. See `amazon.com` for the market price of an (amateur) transcoder.
[36] `zencoder.com`.
[37] `reddit.com/r/Twitch`.
[38] Currently, this problem is no longer an issue, but it was in 2014.

4 The model

The model has three components. Section 4.1 introduces the demand model, section 4.2 the content-provision model, and section 4.3 the congestion model.

4.1 Viewers (demand)

Each period, a representative consumer i decides what channels to watch according to a discrete-choice model.[39] Let $\mathcal{J}$ be the set of all channels that have an account on the platform. Let $\mathcal{J}_t \subset \mathcal{J}$ be the set of online channels at date t, and it always includes the outside option, which is watching nothing. The consumer's outside option has a normalized utility of ε_{i0t}, whereas the payoff for tuning into channel j is $u_{ijt} \equiv \delta_{jt} + \varepsilon_{ijt}$. The consumer solves $\max_{j \in \mathcal{J}_t} u_{ijt}$ and, assuming type 1 extreme value errors, the market-share equation becomes

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt} \equiv \mu_j + \alpha_1 \log \sigma_{jt} + \alpha_2 \tau_{jt} + \boldsymbol{\beta}' \mathbf{x}_{jt} + \xi_{jt}, \tag{1}$$

where μ_j is a fixed effect, σ_{jt} is the noise-to-signal ratio of the channel (a measure of quality of service), τ_{jt} indicates whether channel j's broadcast is transcoded, and ξ_{jt} is unobservable. Channel j characteristics, $\mathbf{x}_{jt}$, include indicators for partnership, for a channel being featured on the platform's homepage, for weekends, and for peak hours (between 6pm and 1am local time); the log of tenure in the platform (as a proxy for quality and for fan-base size); and the log of uptime. Notably, there are no prices. In practice, viewers do not pay cash for entering the platform. Instead, they pay with "eyeballs" as they have to watch ads to view the broadcasts. Moreover, this pricing and ad policy remained constant during the sampled time frame. Therefore, price endogeneity is not a problem, and the constant will absorb any remaining effect.

The assumption of iid T1EV errors causes two implicit consequences: (1) because

[39] See, for example, McFadden (1974) or Berry (1994).

adding an extra channel increases the "inside option" market share, then the model incorporates a positive cross-side network effect[40] from the supply side to the demand side; and (2) because an extra channel decreases the market share of all other channels (via the non-Independence of Irrelevant Alternatives), the model incorporates a negative competition effect on the supply side.

I expect that a high noise-to-signal ratio decreases the demand for a channel, because it will manifest in video buffering, latency, and other anomalies. I also expect a transcoded stream to have more viewers, all else equal. However, I do not expect the partnership status to have a large effect, because the benefits for viewers are negligible once transcoding is taken into account.

An increase in the number of own viewers (ie, downloads) does not increase the noise-to-signal ratio by itself, because streams are downloaded from an edge server, not from the source. Noise can be increased, though not decreased, when the broadcast is downloaded, because of idiosyncrasies in the consumers' connections. I assume those idiosyncratic shocks are absorbed by the T1EV error, ε_{ijt}, which is already integrated out of equation (1). Section 4.3 presents more details about the nature of the noise-to-signal ratio.

I assume the market size is $M = 4.16$ million viewers; see section 4.4 for details. Then $s_{0t}M$ viewers tune out of the platform at any date t. Therefore, $N_t^V = (1 - s_{0t})M$—this expression for N_t^V is known as quasi-demand. On the other side of the market, $\#\mathcal{J}_t$ is the quasi-supply (Rochet and Tirole, 2006).[41]

Identification. In equation (1), the parameters of interest are the coefficients of the noise-to-signal ratio and the transcoder indicator. However, their identification presents some challenges.

[40] See Rochet and Tirole (2006) for a review of two-sided markets.

[41] Market shares are relatively small, so one might be worried about machine precision when the estimation is performed. To mitigate this concern, instead of using market shares, I regressed the log of the absolute number of viewers on the right-hand-side variables plus the market size, with its coefficient constrained to 1. The results are indistinguishable. Indeed, the log of market shares takes values of -15 at least.

First, as an equilibrium outcome, the noise is simultaneously determined with demand. Moreover, noise could be correlated with the error term through selection or through an unobserved component that increases the quality of a stream and at the same time increases its noise. For instance, videos with more movement, which could be more entertaining, require a higher bitrate, which in turn increases its volatility. Yet another concern is measurement error, because σ_{jt} is constructed using the average bitrate as a proxy for the target bitrate.

To mitigate these concerns, the regression includes a fixed effect for channel j, and, moreover, I use two sets of instruments. The first set are the PoP upgrades (see Figure 4), which improved the network infrastructure and decreased noise all over the platform. Moreover, these upgrades should not increase viewership by themselves but only through noise reduction. A second instrument comes from the introduction of the Xbox One broadcasting ability. Upon its introduction, the total number of channels online significantly increased, which in turn increases the upload video bitrate to the platform. The spike in traffic increased the noise-to-signal ratio across the platform. Figure 2 supports the validity of this instrument. Section 4.3 analyzes the mechanics of the noise-to-signal ratio.

The exclusion restrictions would fail if the platform is reacting strategically to viewership instead of, say, following a long-run business plan. I believe the platform is not reacting to the short term. The PoP upgrades and the Xbox One upgrade were planned with anticipation but were released as soon as they were available. For instance, the Xbox One upgrade was unveiled on a Tuesday, which is the second least popular day to broadcast.

Equation (1) presents no random coefficients that allow for flexible substitution patterns. This issue could potentially bias the results. However, the logic that applies to common price elasticities does not translate directly into noise elasticities. A classic problem in a BLP-type model is that own price elasticities are roughly constant (when the model includes log prices), which implies similar markups even for products with different prices. An additional problem is that cross-price elasticities are proportional to market shares.[42] However, back to this setting, channels do not choose their noise-to-signal ratio. Therefore, we have no a priori reason to discard constant own noise elasticities. Finally, to allow for more flexible substitution patterns, I consider a nested logit in Appendix section A. The reason the nested logit is favored over a random-coefficients approach is practical. Because viewers can watch any channel around the world, at any given time, 5,000 "products" are available in market t. And because the data set covers 90 days, more than 13,000 "markets" are included.

Now consider selection. Because noisier channels will have less viewership, those channels would not be online. In fact, the content-provision model of section 4.2 explicitly takes viewership into consideration. In particular, the model assumes that channels begin their broadcast without considering their actual viewership number because it has not been realized yet. Thus, the first few minutes of each broadcast should not be subject to selection. Therefore, to account for the selection rule, I restrict the sample to the first observation of each broadcasting session. Alternatively, Appendix section A considers a Heckman (1979) correction, also based on the content-provision model.[43] Formally, I assume the noise-to-signal ratio is orthogonal to the error, conditional on the instruments and on the selection rule.

Finally, transcoders are guaranteed to partners, who are the most popular channels. Thus, the allocation is correlated with quality. However, transcoders are also allocated according to excess capacity in an unexpected way from the point of view of the channel. The platform follows a conservative strategy to protect itself from sudden surges in partnered broadcasts. Thus, non-partnered, transcoded broadcasts are almost always available. Identification comes from comparing channels before and after receiving a transcoder.

[42]See Berry, Levinsohn and Pakes (1995).

[43]Because a significant proportion of channels have zero market shares, Appendix section A also considers a correction for zero market shares, based on Hortaçsu and Joo (2015).

Formally, I assume transcoders are orthogonal to the error conditional on fixed effects. A possible threat to identification comes from the fact that receiving a transcoder means fewer partnered channels are online, which means less competition. The estimation could be picking up the effect of less competition, instead of the causal effect of a transcoder, and would bias the target coefficient upward. To mitigate this concern, I include controls such as weekend dummies and peak-hour dummies. The results, presented in section 5, are also robust to the inclusion of day-of-week fixed effects and hour-of-day fixed effects.

Although the data set is poor on channel characteristics, it is rich in the time dimension. Notably, channel fixed effects control for important but unobserved attributes such as content, quality, and other horizontal characteristics. The following summarizes the identification assumptions.

Assumption 1. (a) $\log \sigma_{jt} \perp \xi_{jt} | \text{Xbox}_t, \mathbf{PoP}_t$, selection rule, $\mathbf{x}_{jt}$; (b) $\tau_{jt} \perp \xi_{jt} | \mu_j, \mathbf{x}_{jt}$.

4.2 Content providers (supply)

Each period t, each broadcaster $j \in \mathcal{J}$ decides to be online or not.[44] Let u^B_{jt} be the (random) utility of broadcasting at t. The utility of not broadcasting is normalized to zero. Thus, broadcasters solve a static problem: $\max \left\{0, u^B_{jt}\right\}$, where u^B_{jt} depends on the information available at t. In particular, the decision depends on the number of viewers watching j. Note, however, that this information is not available when channel j is offline. For this reason, I separate the broadcasters' decisions by their information sets.

I assume broadcasters solve a static problem that can be divided into two types of decisions: (1) When the channel is offline, the broadcaster has to decide whether to **Turn On** or not. (2) If the broadcaster has decided to be online, this decision will be followed by the decision to **Keep On** or not, in the following periods.

First, when the channel is offline, the channel does not know how many viewers it

[44]In reality, the set $\mathcal{J}$ grows over time as channels sign up in the platform. I will not introduce further notation to account for this fact. Entry (ie, having an account in the platform) is taken as given, but it is controlled for.

could have if it were online. Thus, I model this decision as a function of observables such as the aggregate number of viewers and channels, which proxy as potential market and competition. Importantly, a deciding factor is the promise of a transcoder, which is only guaranteed for partners. Thus, the partner status will be the variable of interest.

Second, given that the channel is already online at date t, the number of viewers is realized and the channel decides whether to Keep On given this information. Thus, being online at $t+1$ is decided at t with the information of period t.

Specifically, let y_{jt} indicate if channel j is online at t. Channel j observes the state of the world and decides whether to Turn On or not at the beginning of period t, according to a utility associated with each decision. Let u_{jt}^{IO} be the utility of Turning On, and let 0 be the utility of staying offline. Thus, channel j Turns On iff $u_{jt}^{IO} > 0$. That is, I assume

$$y_{jt} = 1 \,|\, y_{j,t-1} = 0 \iff u_{jt}^{IO} > 0,$$

$$u_{jt}^{IO} \equiv \mu_j^{IO} + \alpha_1^{IO} \mathbb{1}\{\text{Partner}_{jt}\} + \alpha_2^{IO} \mathbb{1}\{\text{Xbox}_t \times \text{Gaming}_j\} + \boldsymbol{\beta}^{IO\prime} \mathbf{x}_{jt}^{IO} + \varepsilon_{jt}^{IO}, \tag{2}$$

where μ_j^{IO} is a fixed effect; $\mathbb{1}\{\text{Partner}_{jt}\}$ is an indicator for partnership; and $\mathbb{1}\{\text{Xbox}_t \times \text{Gaming}_{jt}\}$ is an interaction of the Xbox One broadcasting availability and the gaming category of channel j. The controls $\mathbf{x}_{jt}^{IO}$ include information available before the transmission starts: the log of viewers per channel, $\log\left(N_t^V / N_t^B\right)$, as a measure of potential viewership; indicators for weekend and for peak hours (defined as 6pm to 1am local time); log of downtime (time spent offline); and as a proxy for experience or for developing a fan base, the log of tenure.

All else equal, one would expect that partners are more willing to broadcast, because they know they will have a transcoder, which will increase their viewership. The Xbox shock should affect only gaming channels, which account for 95% of the population.

Given that the channel decided to Turn On at t, the viewers of channel j are realized and observed by j at t. With this information, at the end of period t, the channel decides whether to Keep On or to Turn Off at $t+1$. Let u^B_{jt} be the utility of Keeping On, and let 0 be the utility of Turning Off. Thus, channel j Keeps On iff $u^B_{jt} > 0$:

$$y_{j,t+1} = 1 \,|\, y_{jt} = 1 \iff u^B_{jt} > 0,$$

$$u^B_{jt} \equiv \mu^B_j + \alpha^B_1 \log n^{V,\text{Partner}}_{jt} + \alpha^B_2 \log n^{V,\text{NonPartner}}_{jt} + \alpha^B_3 \mathbb{1}\{\text{Partner}_{jt}\} + \boldsymbol{\beta}^{B\prime}\mathbf{x}^B_{jt} + \varepsilon^B_{jt}. \quad (3)$$

where μ^B_j is a fixed effect, n^V_{jt} is the number of viewers watching j at t, and it is interacted with the partnership dummy. The controls, $\mathbf{x}^B_{jt}$, include $\mathbb{1}\{\text{Xbox}_t \times \text{Gaming}_j\}$, indicators for weekends and for peak hours (defined as 6pm to 1am local time), the log of uptime, and the log of tenure.

The effect of own viewership, n^V_{jt}, is separated between partners and non-partners to allow different marginal valuations. Partners have more viewers on average, which should drive down the marginal valuation of a viewer. However, partners also receive a share of ad revenues and subscription revenues, both of which are proportional to their viewership numbers. For these reasons, partners are expected to have different intercepts and coefficients.

At this point, the likelihood of observing a given history of a channel can be written. Let T^s_j define the set of periods when channel j had its s-th broadcasting session. That is, $T^s_j \equiv \left\{t^s_{1j}, \ldots, t^s_{end,j}\right\}$, where the channel was offline at time $t^s_{1j} - 1$, Turned On at t^s_{1j} and was online until $t^s_{end,j}$, when it decided to Turn Off. That is, the typical s-th session is $\left\{y_{jt}\right\}_{t \in T^s_j} = \{1, 1, \ldots, 1, 1, 0\}$. Any channel j could have S_j of such broadcasting sessions. The likelihood of session s is:

$$\begin{aligned} L^s_j\left(\boldsymbol{\theta}^{IO}_0, \boldsymbol{\theta}^B_0\right) = P\left[y_{jt^s_{1j}} = 1 \,|\, y_{j,t^s_{1j}-1} = 0, \boldsymbol{\theta}^{IO}_0\right] \left[\prod_{t=t^s_{1j}}^{t^s_{end,j}-2} P\left[y_{j,t+1} = 1 \,|\, y_{jt} = 1, \boldsymbol{\theta}^B_0\right]\right] \\ \times P\left[y_{j,t^s_{end,j}} = 0 \,|\, y_{j,t^s_{end,j}-1} = 1, \boldsymbol{\theta}^B_0\right], \end{aligned}$$

where $\boldsymbol{\theta}_0^{IO}$ and $\boldsymbol{\theta}_0^{B}$ are the vectors of parameters from the decisions of Turning On and Keeping On.

This likelihood is conditional on the fact that (by definition) the channel was offline before the session starts. To wrap up the likelihood of the data, define $T_j^{\text{off}} \equiv T \setminus \bigcup_{s=1}^{S_j} T_j^s$ as the inter-session periods when the channel was offline, and where T is the set of periods in the data. Then,

$$L_j\left(\boldsymbol{\theta}_0^{IO} | y_{j0}\right) = \prod_{t \in T_j^{\text{off}}} P\left[y_{jt} = 0 \,|\, y_{j,t-1} = 0, \boldsymbol{\theta}_0^{IO}\right]$$

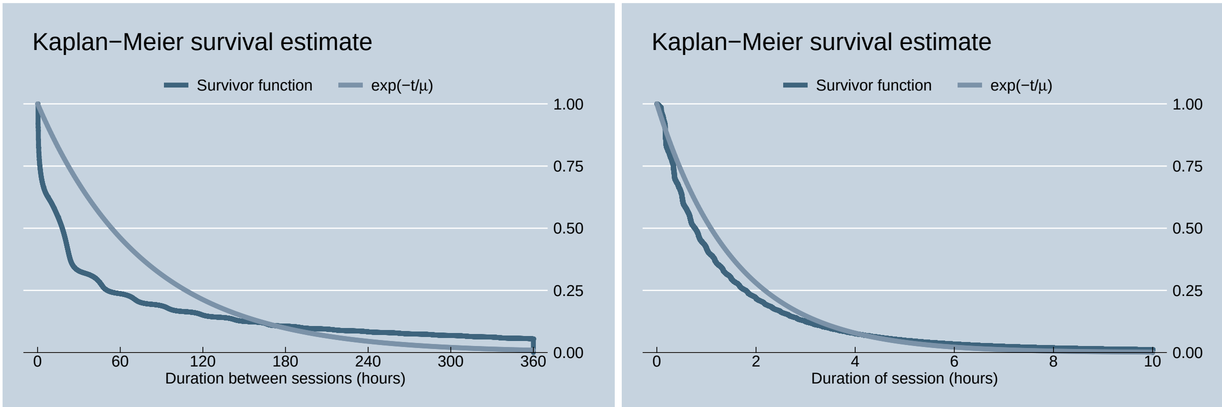
is the likelihood of observing channel j being offline in the inter-session periods, conditional on the starting value y_{j0}.

Finally, the likelihood of the data can then be written by considering all sessions for all channels:

$$L\left(\boldsymbol{\theta}_0^{IO}, \boldsymbol{\theta}_0^{B}\right) = \prod_{j \in \mathcal{J}} L_j\left(\boldsymbol{\theta}_0^{IO} | y_{j0}\right) \prod_{s=1}^{S_j} L_j^s\left(\boldsymbol{\theta}_0^{IO}, \boldsymbol{\theta}_0^{B}\right). \tag{4}$$

A simplifying assumption is implicit in (4): I assume equations (2) and (3) are independent of each other so that $\boldsymbol{\theta}_0^{IO}$ and $\boldsymbol{\theta}_0^{B}$ do not share parameters. That is, I assume the likelihood is separable and hence the parameters can be obtained through two separate maximization programs. Thus, the working assumption is that the errors from equations (2) and (3) are independent.

The potential problem with this assumption is that a shock that pushes a channel to Turn On could be correlated with the shocks a channel experiences once it is online. That is, the errors could be correlated across time. I do not consider this issue to be a major setback, because I include individual effects on both equations and I allow for panels in the likelihood. I also include tenure, uptime, and downtime data which allow me to control for unobserved dynamics. Moreover, the likelihood conditions on the outcome of the previous period, and the presence of autocorrelation does not bias the estimates.



Note: Downtime and uptime comparison. Kaplan-Meier survival estimates and a fitted exponential curve.

Figure 6: Comparison of downtime and uptime

Because the problems are separable, the decision to Turn On can be modeled as a binary decision, and the $\boldsymbol{\theta}_0^{IO}$ parameters can be estimated with a stacked probit or logit. The same applies to the $\boldsymbol{\theta}_0^{B}$ parameters in the decision to Keep On. I use a conditional logit in which the fixed effects drop out.

In particular, this model is consistent with a survival model with exponential decay, in which each period the broadcaster can change state. Figure 6 presents non-parametric Kaplan-Meier survival estimates for downtime, where Turning On is the change of state, and for uptime, where Turning Off is the change of state. The figure also shows a simple exponential curve, with no covariates, approximating the Kaplan-Meier curve. Allowing for a more flexible model improves the fit.

Identification. Consider first the problem of estimating $\boldsymbol{\theta}_0^{IO}$ in (2). The parameter of interest is α_1^{IO}, the effect of partnership in the decision to Turn On. Being a partner is potentially endogenous, because partners are expected to have a higher quality and to be the most popular channels. Broadcasters who are more able might derive more utility from broadcasting.

Partnerships are never lost, but they are gained. In the sample, I observe more than 1,600 partners of whom more than 600 acquired partnerships. To be a partner, broadcasters must apply through the platform's website and be approved. Broadcasters know they have applied for partnership but do not know if and when they will be approved. The platform approves partnerships on a rolling basis. From the broadcaster's point of view, partnership comes unexpectedly. Thus, to identify the causal effect of partnership, I exploit within-panel differences of channels before and after they become partners. Formally, I assume partnership is exogenous, conditional on the fixed effect.

Consider now the problem of estimating $\boldsymbol{\theta}_0^B$ in (3). The same concerns about the effect of partnership apply here too. Thus, I also exploit within differences to identify the effect of partnership. A different problem arises about the exogeneity of the number of viewers, n_{jt}^V. Viewership could be correlated with the error term for a number of reasons. First, unobserved quality could have a positive effect on the utility of the broadcaster. Second, dynamic considerations, such as inertia of viewers, could imply lagged variables or differences, such as $\Delta \log n_{jt}^V$, are omitted variables. To identify the effects of n_{jt}^V, I use the demand-side model to construct a control function.[45] Specifically, in the first stage, I estimate

$$\log n_{jt}^V = \mu_j^F + \alpha_1^F \log \sigma_{jt} + \alpha_2^F \tau_{jt} + \boldsymbol{\beta}^{F\prime} \mathbf{x}_{jt}^F + \eta_{jt}, \tag{1$'$}$$

with η_{jt} and ε_{jt}^B correlated with each other, and where, notably, $\log \sigma_{jt}$, τ_{jt}, and a dummy for being featured are excluded from equation (3). Then I construct $\widehat{\eta}_{jt}$ and include it as an additional regressor in (3), interacted with the partnership dummy: $\widehat{\eta}_{jt} \times \mathbb{1}\{\text{Partner}_{jt}\}$ and $\widehat{\eta}_{jt} \times \mathbb{1}\{\text{NonPartner}_{jt}\}$. Section B.1 provides details on how to correct the scale of the coefficients in the logit with control functions.

In estimating equations (2) and (3), I use a conditional logit model, which drops out the fixed effects and avoids an incidental parameters problem. The fixed effects allow me to control for unobserved but important attributes such as content, quality, or ability, which may be correlated with partnership or viewership.

By assuming iid errors, I implicitly assume dynamic considerations are not important. This holds if, for example, channels can predict their viewers in the next 10 minutes

[45]See Petrin and Train (2010).

by observing their current number of viewers. To control for unobserved viewership growth or lagged variables, I include uptime and tenure. The following summarizes the assumptions.

Assumption 2. (a) $\varepsilon^{IO}_{jt}, \varepsilon^{B}_{jt}$ are iid logistic; (b) $\mathbb{1}\{\text{ Partner }_{jt}\} \perp \varepsilon^{IO}_{jt} | \mu^{IO}_{j}, \mathbf{x}^{IO}_{jt}$; (c) $\mathbb{1}\{\text{Partner}_{jt}\} \perp \varepsilon^{B}_{jt} | \mu^{B}_{j}, \mathbf{x}^{B}_{jt}$; (d) the error ε^{B}_{jt} can be decomposed into the part that can be explained by a general function of η_{jt} and a residual: $\varepsilon^{B}_{jt} \equiv \text{CF}(\eta) + \widetilde{\varepsilon}^{B}_{jt}$; (e) $\log n^{V}_{jt} \perp \varepsilon^{B}_{jt} | \widehat{\eta}_{jt}$.

4.3 Noise-to-signal ratio

To close the model, this section describes how congestion is generated.

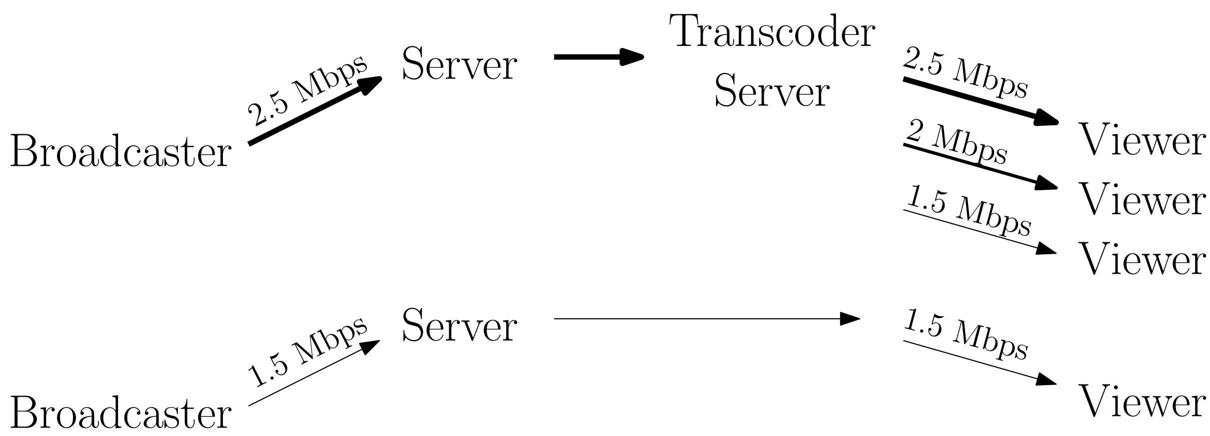
Let b_{jt} stand for the (upload) video bitrate of channel j at date t, and let $\overline{b}_j$ be its mean over t. Then the noise-to-signal ratio[46] σ_{jt} is defined as

$$\sigma_{jt} \equiv \frac{|b_{jt} - \overline{b}_j|}{\overline{b}_j}.$$

If σ_{jt} increases, the stream's quality starts to decrease: the stream may buffer, or the video may get "choppy" due to the loss of data packets and the extra stress on the hardware—video encoding and decoding is CPU-intensive, or the stream may present other anomalies such as latency. This quality measure is appropriate for the platform, because Twitch requires broadcasters to stream using a constant bitrate protocol. Thus, in principle, b_{jt} should be constant, and $\overline{b}_j$ is a consistent estimate of the target bitrate.

The "production" of noise will be composed of two factors: an individual factor and an aggregate factor. The individual factor represents the broadcaster j's private internet speed, the overall quality of the connection, and shocks beyond the platform's influence. The aggregate factor is common to all broadcasters and is rooted in the platform's physical capacities. This aggregate factor is a function of aggregate participation and the

[46]See Shivaldova, Winkelbauer and Mecklenbrauker (2014) for an example of a common parametrization in the engineering sciences.



Note: An (abstracted) example of a broadcast's path from upload (↗) to download (↘). If the broadcast is transcoded, it reaches a transcode server which distributes the data at lower bitrates. If it is not, the broadcast is mirrored (→) at source speed.

FIGURE 7: Data flow with transcoding v without

manifestation of congestion externalities.

Not all traffic has the same effects on congestion. Specifically, uploads are processed separately from downloads. The upload reaches a server, from which it is mirrored through the platform's network, and then distributed to edge servers, from which viewers download the stream. See Figure 7 for reference. Noise can be generated while uploading or downloading. Noise on the upload side is caused by traffic into the upload servers and will be transmitted to all the viewers watching the broadcast. However, the noise that is generated while downloading the stream is idiosyncratic to each consumer and does not affect the upload side of the network. Only the aggregate uploaded video bitrate will be considered as a source of congestion for my measure of the noise-to-signal ratio.

I assume

$$\log \sigma_{jt} = \mu_j^{\sigma} + \alpha_1^{\sigma} \log B_t^{up} + \boldsymbol{\beta}^{\sigma\prime} \mathbf{PoP}_t + \varepsilon_{jt}^{\sigma}, \tag{5}$$

where B_t^{up} is the total video bitrate uploaded to the platform's servers at date t and $\mathbf{PoP}_t$ is a vector of four indicators for the point of presence upgrades.

Identification. Although individual channels are too small to have an effect on aggregate variables, the decision to be online depends on the noise-to-signal ratio. If noise is too high, the channel will turn off and it will not be observed. We thus have a selection problem. Moreover, if every channel is selected in this way, we would only observe low levels of noise, even as more channels turn on. That is, as B_t^{up} increases, channels with higher noise-to-signal ratios drop out. I use the Xbox One shock as an instrument for B_t^{up} because the shock increased the number of broadcasters, which increased the upload traffic.

To solve the selection problem, I implement corrections based on the content-provision model of section 4.2. These corrections parallel the ones used in the demand-side model to solve the same selection problem. Specifically, channels begin their broadcast without considering their actual viewership number, because it has not been realized yet. Thus, the first few minutes of each broadcast should not be subject to selection. Additionally, Appendix section B considers a Heckman (1979) two-step correction in which the first stage comes from the decision to Keep On. This decision uses the realizations of the number of viewers to determine the probability that the channels is online the next period.

Finally, I include fixed effects in this estimation because no cross-sectional variation exists over B_t^{up}. The fixed effects also control for unobserved channel characteristics such as location, internet connection, or hardware.

4.4 Market size

In February 2014, Twitch.tv had about 45 million unique viewers each month. By the end of 2014, the number had reached 100 million. Assuming linear growth, by the end of April 2014, the platform would have grown to around 50 million viewers. Also, the platform reports that each viewer watches an average of 106 minutes per day. Thus, I assume viewers have 120 minutes available per day to watch videos. Then, assuming a uniform distribution over the day, which is a reasonable approximation to the data (Pires and Simon, 2015), I define the market size as 4.16 million viewers available at any given moment.[47] The results are not sensitive to this assumption.

4.5 Discussion

I consider a monopolistic platform and a static framework. Section 6 discusses the platform's incentives to invest in infrastructure, but I abstract away from the platform's problem and consider its strategy constant over the sample period. As argued above, we have reasons to believe the platform's planning horizon is in the long run. Thus, this paper is best interpreted as a short-run analysis.

Note here that this is an equilibrium model. Indeed, suppose that channel 1 decides to be online at t. Then the consideration set $\mathcal{J}_t$ changes for every consumer. Then the demand for every channel $j \neq 1$ changes, as well as the noise-to-signal ratio for every channel. Thus, the incentives of every channel have changed, and some channels may now decide to be offline. Again, demand and noise would change for every channel.

Each dot in Figure 3 is an equilibrium outcome. I expect that network effects push the platform to grow, whereas the congestion externality pulls it back. For example, upon a shock that increases content provision, demand increases more than expected due to a network effect, which, in turn, causes feedback over the supply side, and more content providers begin supplying. However, congestion decreases the quality of the streams, which decreases demand. The network effect again pushes supply down more than otherwise expected.

[47] If 50m viewers are available in a day, each watching 2 hours out of 24, then 4.16m $\approx$ 50m $\times$ 2/24 = 4.1$\overline{6}$. I use the exact number, 4,166,666, in the actual estimation.

5 Estimation results

5.1 Demand-side results

Table 5 presents the results of the demand estimation. Column 1 shows an OLS, and column 2 introduces fixed effects. Because transcoders are allocated most commonly to partners, one would overestimate the benefit of transcoders. Conditional on fixed effects, transcoders are allocated (as good as) randomly. Column 3 presents a 2SLS with both types of instruments: the Xbox One shock and the PoP upgrades. Column 4 adds fixed effects, whereas column 5 adds relevant controls. Finally, column 6 addresses the selection problem by restricting the sample to the first observation of each broadcasting session. The logic is that immediately after Turning On, a channel does not know its noise-to-signal ratio. The effects remain qualitatively similar. By using a subsample, we lose power and the standard errors increase. Appendix section A considers a different selection-correction strategy based on Heckman (1979). The results are robust.

As expected, an increase in the noise-to-signal ratio decreases demand. Because market shares are small, $(1 - s_{jt}) \approx 1$, the estimated coefficient of $\log \sigma_{jt}$ can be interpreted as the elasticity of own noise-to-signal ratio. In other words, if σ_{jt} increases by 1%, demand decreases approximately .2%. This number is economically significant because fluctuations of σ_{jt} in the order of 50% are common. Moreover, when paired with the effect of B_t^{up}, which can also increase by 50% in a single day, the congestion externality looms large as traffic affects *all* channels. Details follow at the end of this section.

Also, as expected, the transcoder increases demand in a significant way, approximately by 25%. This effect is robust when controlling for partnership, features, tenure, weekends, peak hours and uptime. These covariates, as well as the fixed effect, help control for unobserved quality.

As expected, conditional on having a transcoder, the partnership status of a channel does not increase viewership, because the other perks of partnered channels, such as special chat interactions and "virtual badges," offer minimal benefits to viewers.

The first stages are reported on Table 9 in Appendix section A (see also the results on the noise-to-signal ratio estimation, section 5.4). The relevant columns are 1 through 4. In column 4, when selection is addressed, we can see negative coefficients on the PoP upgrades. Moreover, the coefficients loosely align with the relative importance of these upgrades. Although precise numbers were not released, the first upgrade was the largest, followed by the third, followed by the second. No information is available on the relative size of the fourth upgrade.[48]

5.2 Supply-side results: Turning On

Each channel j is too small to have an impact on aggregate variables. That said, a platform-wide shock might occur that could drive all channels and viewers in or out the platform. In such a case, the error and the aggregate variables could be correlated. To control for these type of shocks, I include the log of viewers per channel, $\log\left(N_t^V/N_t^B\right)$; indicators for weekend, for peak hours; the log of time being offline (evidently necessary from Figure 6); and the log of tenure.

In practice, to estimate the parameters, I use a random sample for computational concerns: the data contain more than 1 million channels and more than 12,000 10-minute periods in the data. The results do not seem sensitive for the selected sample size.

Results are in Table 6. The first two columns show a logit model. The partnership status is expected to be upwardly biased because it is correlated with unobserved channel quality. The next two columns show a conditional logit, in which channel fixed effects drop out. Conditional on fixed effects, the partnership coefficient has a causal interpretation: the partners' probability of Turning On is about 50% higher than that of non-partners. The baseline probability of Turning On is .07% at any given 10-minute period in time. Table 13 in Appendix section B presents the marginal effects and a linear probability model. The

[48] `blog.twitch.tv`.

results are robust after including time fixed effects.

Partners are more likely to Turn On for two reasons. First, the promise of a transcoder increases expected viewership. Second, viewers bring about a monetary compensation for partners. These two effects are captured by the same parameter because a non-partnered channel is unable to predict if, upon Turning On, it will have a transcoder.

Indeed, the only variation in partnership status comes from would-be partners that have applied to the program and are accepted. With this sole source of exogenous variation comes the caveat of extrapolating the effect of partnership out of sample. That is, I assume a change in partnership would affect any channel in this way.

As expected from Figure 6, participation is less likely with downtime. However, channels Turn On when there are more viewers per channel, in weekends, and in peak hours. Their probability of Turning On decreases with tenure.

5.3 Supply-side results: Keeping On

The results for the decision to Keep On are in Table 7. One would expect more able broadcasters to derive higher utility from broadcasting and, at the same time, for their viewership and partnership status to be positively correlated with unobserved ability. Column 1 shows a baseline logit. Column 2 adds relevant controls. Column 3 is a conditional logit in which fixed effects drop out. Conditional on fixed effects, the effect of partnership is identified. To identify the effects of partnership, column 4 adds two control functions. The first stage, described in equation (1′), yields residuals $\widehat{\eta}$. I then include their interactions with a partnership dummy as control functions.[49]

As expected, partners stay online for longer. Given that they are online, the probability of Keeping On increases 33 percentage points from a baseline of 63% for non-partners. That is, combined with the results from the previous section, in both their extensive and

[49]Column 4 already shows the coefficients in the original scale, as the decomposition of the original logistic error, $\varepsilon \equiv \lambda_1 \eta \times \mathbb{1}\{\text{Partner}\} + \lambda_2 \eta \times \mathbb{1}\{\text{NonPartner}\} + \widetilde{\varepsilon}$, induces a logit with a lower variance: $\mathbb{V}[\widetilde{\varepsilon}] \leq \mathbb{V}[\varepsilon]$. See Appendix section B.1.

intensive margins, partners broadcasts more.

Unexpectedly, the coefficient on viewership is lower for partners than for non-partners. Partners receive compensation proportional to their audience. The conversion rate of viewers into subscriptions might drop rapidly in the absolute number of viewers, and hence the drop in marginal valuation. Moreover, conditional on being a partner, broadcasters are already likely to Keep On streaming, regardless of viewership. Finally, partners have significantly higher viewership numbers, which do not change meaningfully over time. This low variation translates into loss of power, as shown in Table 14 in Appendix section B, where I restrict the sample to channels that were partnered at some point.

The first stage is shown in Table 15 in Appendix section B. In line with the demand-side results in section 5.1, viewership decreases with noise and increases with transcoding and features.

5.4 Congestion results

Table 8 shows the estimation results of the noise equation (5). Column 1 shows an OLS. Although channels are too small to generate reverse causality, as the aggregate traffic goes up, the noisier channels might drop first. This effect may confound the coefficient of B_t^{up}. Column 2 instruments B_t^{up} with an Xbox One dummy that indicates when the new technology was available. The idiosyncrasy of internet connections suggests fixed effects. Moreover, no cross-sectional variation exists in B_t^{up}. Column 3 includes fixed effects. Finally, column 4 solves the selection problem by restricting the sample to the first observation of each broadcasting session. Again, in line with the Turning On decision of equation (2), the first few minutes, when the noise-to-signal ratio is yet to be known, should not be subject to selection.

The results are robust to a selection correction based on Heckman (1979) and to the inclusion of a time trend. See Table 16 in Appendix section C. The first stages can be found in Table 17. Overall, the Xbox One shock has a positive effect on the aggregate, upload video bitrate.

The results imply increasing the aggregate, upload video bitrate by 1% will increase the noise-to-signal ratio of *every* channel by about .4%. In a single day, B_t^{up} can fluctuate by 50%. An increase of 50% in B_t^{up} would translate into an increase of 20% in the noise-to-signal ratio of every channel. According to the demand-side results, increasing the noise by 20% will decrease demand by about 4%. In sum, the partial-equilibrium effect of an upload traffic increase of 50% is an aggregate demand decrease of about 4%.

6 Counterfactuals

The point of this paper is to simulate a net-neutrality regime on the platform and assess the equilibrium responses. For this exercise, this platform has two relevant levers to be pulled, namely, transcoders and partnerships. These two levers are the ways the platform gives priority to some channels over others. A net-neutrality regime would require the platform to treat all channels the same. Thus, I ask what happens when the allocation of transcoders and partnerships is changed in a non-discriminatory fashion.[50]

The counterfactuals will consist of three regimes: neutral, preferential, and preferential with a rent-extractive platform. In the neutral regime, partnerships will be allocated randomly through the population—partnership implies transcoding. In the preferential regime, channels will be ranked by their quality, using the estimated fixed effects from the demand-side model in section 4.1 as proxies for quality. Then partnerships will be given to the highest-ranked channels. Finally, in the preferential regime with a rent-extractive platform, partnerships will also be given to the highest-ranked channels, but the platform will charge them just enough to offset the benefits accrued from partnership. Section 6.1 provides further details about the implementation of the regime with rent-extraction.

For a given regime, label as c the counterfactual of having enough transcoders for c% of

[50]A third potential lever is the features on the website's homepage. I will shut down the features to focus on the effect of partnership and transcoding. Moreover, the number of features is restricted by the length of

the population. For example, the platform could have resources to guarantee transcoders to $c = 1\%$ of the population. I consider $c = \{0, 1, \ldots, 100\}$.

For each regime and for each counterfactual c, the algorithm to find the equilibrium outcome is as follows. Take a sample $J \subset \mathcal{J}$ of channels. For each simulation $s = 1, \ldots, S$:

1. Allocate partnerships to c% of J according to the given regime.

2. Given observed probabilities of being online, simulate initial online/offline status, $\{y^s_{j0}\}$. Simulate relevant covariates taken from observed distributions. Simulate structural, random shocks ξ^s_j, $\varepsilon^{IO,s}_j$, $\varepsilon^{B,s}_j$ and $\varepsilon^{\sigma,s}_j$, $\forall j \in J$.

   (a) Given 2 of {Demand, Supply, Congestion}, predict the third one and update. That is, given $\{y^s_{j0}, \sigma^s_j\}$, shocks, and covariates, predict $\{n^{V,s}_j\}$. Then, given $\{n^{V,s}_j, \sigma^s_j\}$, predict $\{y^s_j\}$. Then, given $\{n^{V,s}_j, y^s_j\}$, predict $\{\sigma^s_j\}$.

   (b) Repeat until convergence.

For each s, I calculate the consumer surplus, total channels online, total viewers online, the average noise-to-signal ratio, and the average quality of online channels. I perform $S = 5,000$ simulations per counterfactual c, and I draw a random sample of $\#J = 50,000$ channels from the population for the simulations. I assume that each channel has a weight of 20 for the total population size to be 1,000,000.[51]

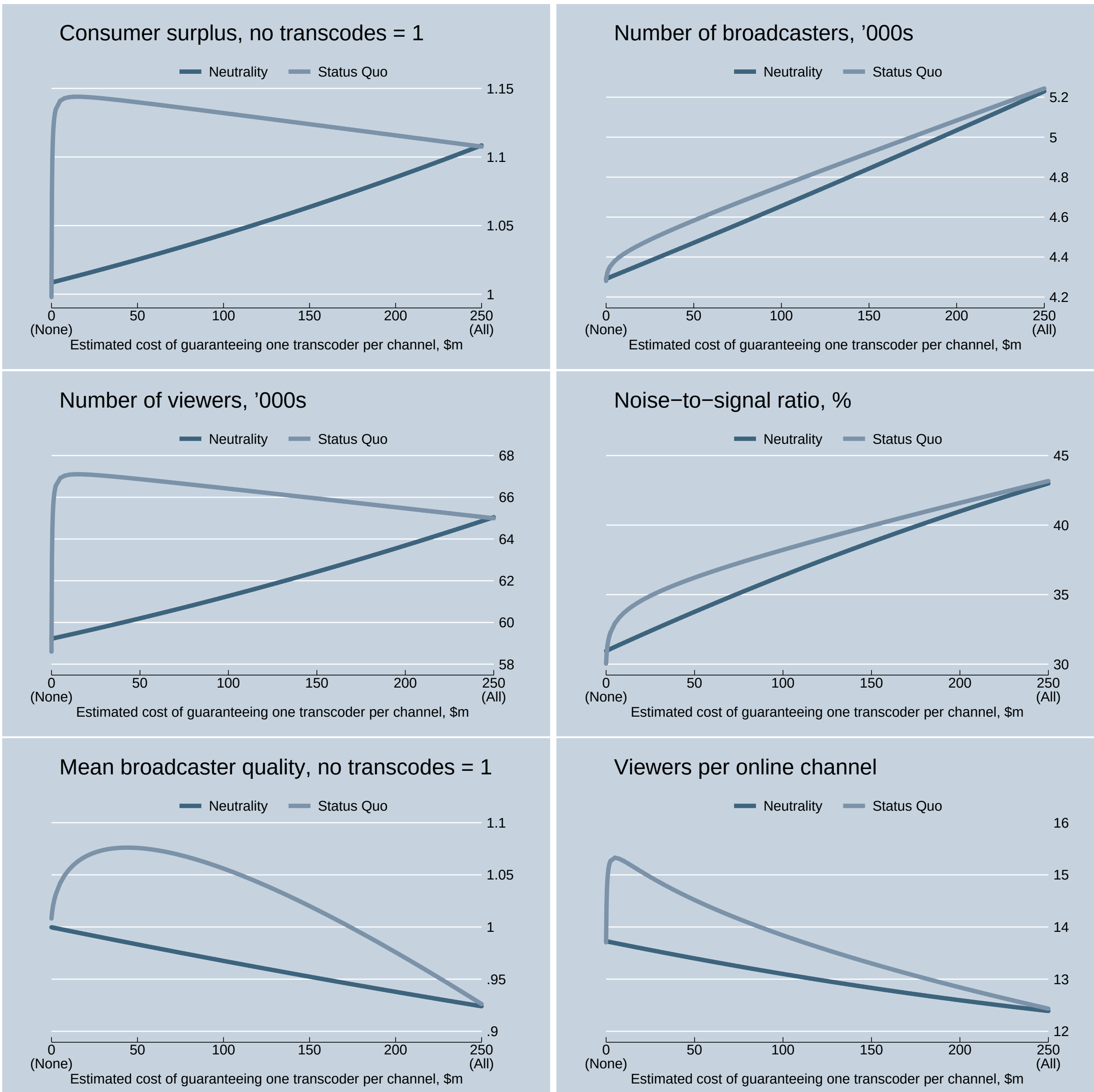
An outstanding issue is the equilibrium selection. Intuitively, the model is essentially linear and well behaved. Because partial derivatives do not change signs in the relevant variables, the model should have a unique equilibrium.

Figure 8 shows the results for the baseline comparison between the neutral and the preferential regime. The x-axis shows the estimated cost of purchasing enough transcoders to guarantee one for c% of the population, up to \$250 million, the cost of acquiring 1,000,000

the website.

[51]The immediate caveat is the out-of-sample prediction as we go toward $c = 100$. However, as a mitigating factor, the model considers individual decisions, and the counterfactual simulations modify the structural parameters individually.

transcoders. For clarity, each line shows a fractional polynomial fit that runs through 101 points; each point represents a different counterfactual $c \in \{0, 1, \ldots, 100\}$. The status quo, which is approximately $c = 1$, is marked as a red line.[52]



Note: Fractional polynomial curves fitted over averages of counterfactual simulations. Averages computed from 5,000 equilibrium simulations (see also Appendix section D). The red line marks the status quo, which is the preferential regime with enough transcoders for about 1% of the population.

FIGURE 8: Neutral v Preferential regimes

[52] The scatter plot with confidence intervals underlying each graph is shown in Appendix section D.

In this benchmark case, neutrality yields a lower consumer surplus by construction. Because transcoders are scarce, their random allocation must yield a less efficient outcome than an allocation that favors the most proficient channels. However, this benchmark is useful to quantify the 5% drop in consumer surplus caused by an immediate shift to neutrality.[53]

Under the preferential regime, the first transcoders are allocated to the channels with higher quality. Here are the largest gains, because those channels make the best use of transcoders by reaching more viewers. These initial gains offset the increase in the noise-to-signal ratio from an increase in traffic. However, as the platform expands its transcoding infrastructure, the relative efficiency gains shrink because the marginal channel's quality starts to decrease while the noise-to-signal ratio keeps increasing. In effect, the congestion externality increases the opportunity cost of transcoders.

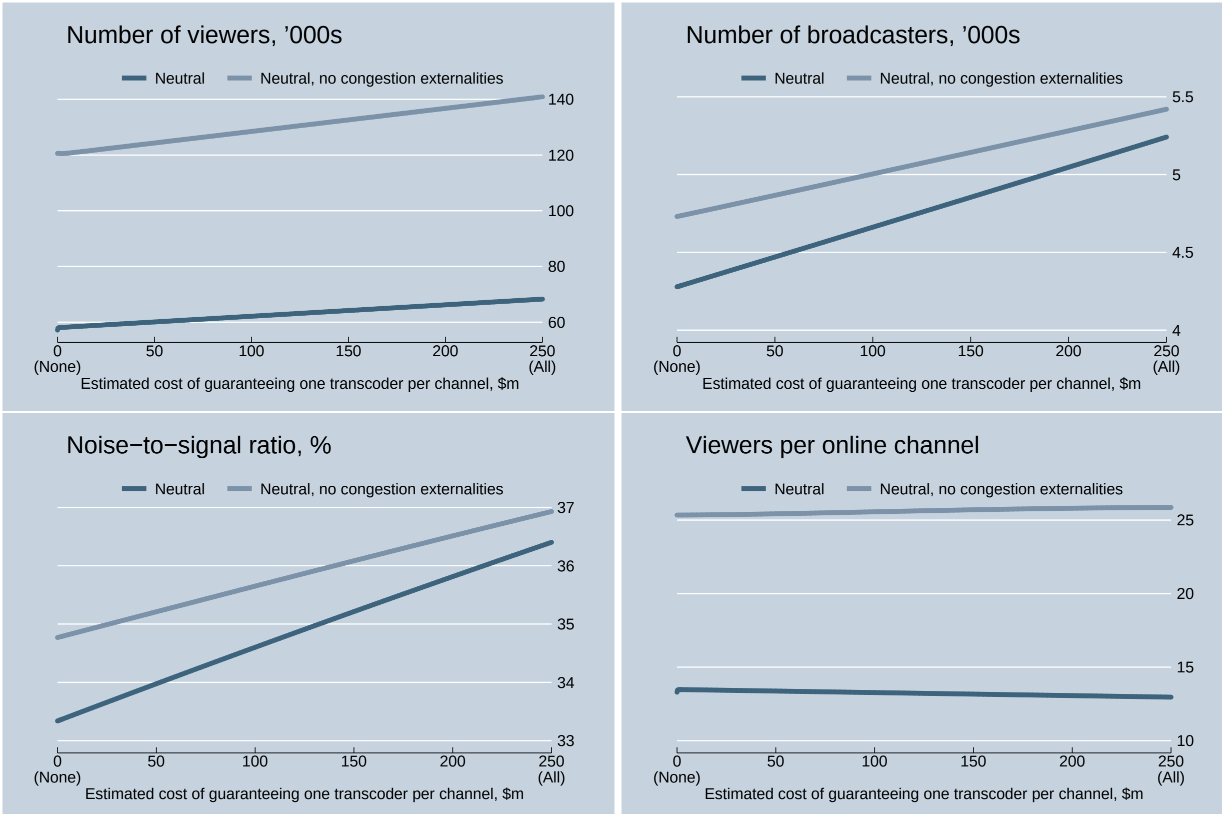
Indeed, congestion externalities play a major role in the growth of the platform. Figure 9 compares the neutral regime with and without congestion externalities. When the externality is turned off, the noise-to-signal ratio is assumed to be constant and equal to its fifth percentile, regardless of aggregate traffic. I still compute the noise-to-signal ratio even if viewers ignore it. The exercise shows that if not for congestion, participation in the platform would be significantly higher.

6.1 Rent-extractive platform v neutrality

Proponents of net neutrality argue that platforms will use their bargaining power to extract rents from content providers. Critics argue that competition and market forces will limit the platform's rent-extractive policies. In this counterfactual experiment, arguably the worst-case scenario for content providers, I explore a monopolistic platform with all the bargaining power.

For this counterfactual, I assume the platform has perfect information and I assume

[53] Consumer surplus is measured in utils and is closely related to demanded quantity, which in this case,



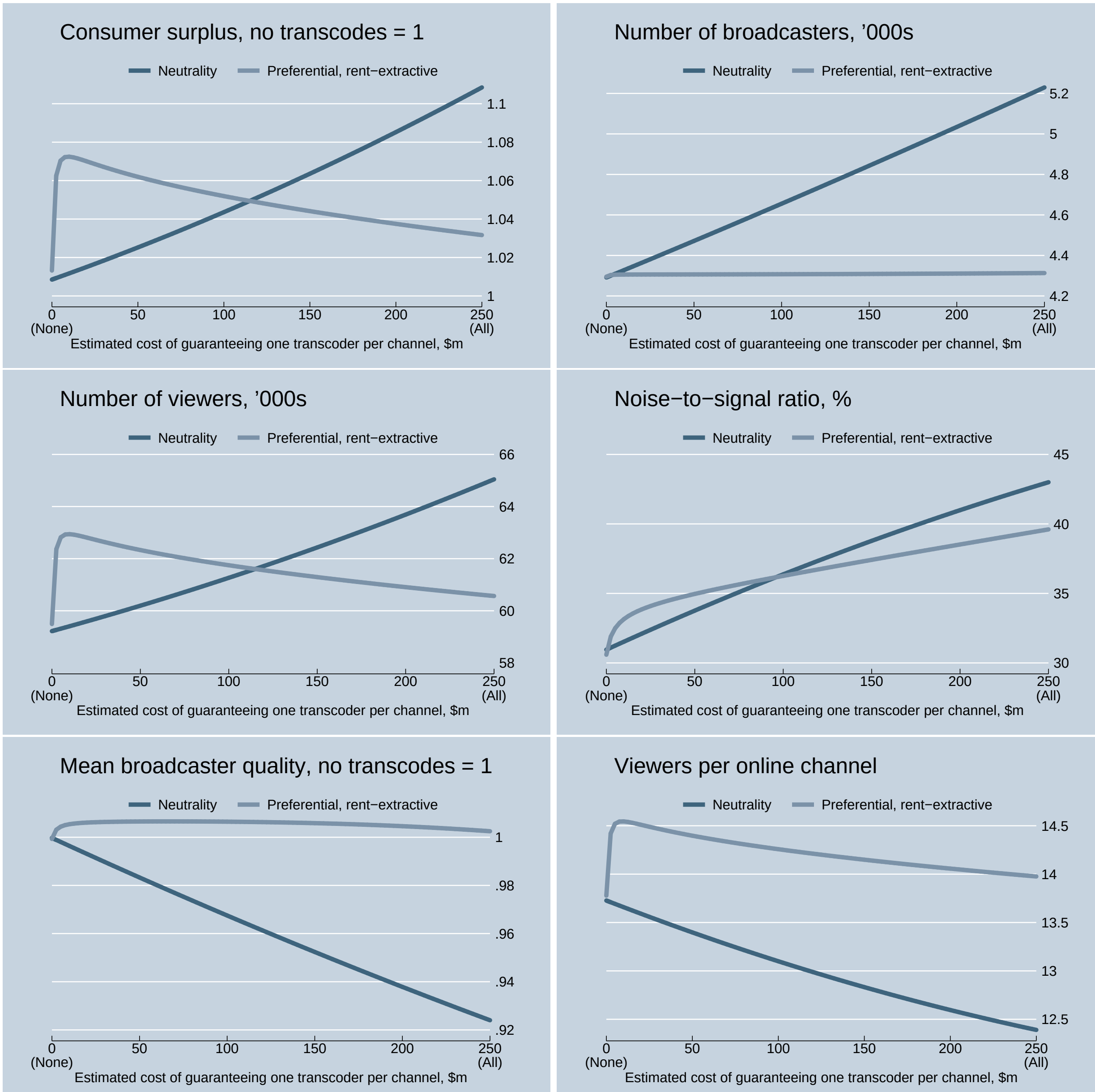
Note: Fractional polynomial curves fitted over averages of counterfactual simulations. Averages computed from 5,000 equilibrium simulations. When the congestion externality is turned off, the effect of the noise-to-signal ratio on demand is assumed to be constant.

FIGURE 9: Neutral regime with and without congestion externalities

that the platform charges channels just enough to leave them indifferent between being partners and not. As a reminder, from the demand-side equation (1), a transcoder increases the log market share of channel j in α_2. From the decision to Turn On in equation (2), the coefficient of partnership is α_1^{IO}. Finally, from the decision to Keep On in equation (3), the coefficient of partnership is α_3^B while the coefficient of log viewers for a partner is α_1^B. Specifically, the rent-extractive platform will charge partners: $\widehat{\alpha}_1^{IO}$ utils when channels decide to Turn On; and $\widehat{\alpha}_1^B\widehat{\alpha}_2 + \widehat{\alpha}_3^B$ utils when they decide to Keep On.

As expected, neutrality encourages content provision. However, in equilibrium, the noise-to-signal ratio also increases while the marginal quality of the extra channel decreases. Initially, the congestion externality looms larger, as we can see from the decrease

is the total number of viewers.



Note: Fractional polynomial curves fitted over averages of counterfactual simulations. Averages computed from 5,000 equilibrium simulations (see also Appendix Section D).

FIGURE 10: Neutral regime v preferential regime with rent-extractive platform

in viewers per channel. However, as the transcoding infrastructure expands, the total number of viewers reaches a plateau in the preferential regime. When transcoders become less scarce, the need for an efficient use decreases while the relative gains from more content provision increase.

Finally, consider the platform’s incentives. Twitch receives revenue mainly from ads and secondly from subscriptions. Both sources of revenue are proportional to the total number of viewers in the platform. As Figure 10 shows, initially, viewers increase quickly with transcoders but slowly afterwards. Therefore, the platform may try to increase its investment in infrastructure at the beginning, but the marginal investment converts less revenues relatively early into the expansion. Therefore, a net-neutrality regime will slow down the investment of small platforms but will increase that of larger ones.

6.2 Heterogeneity

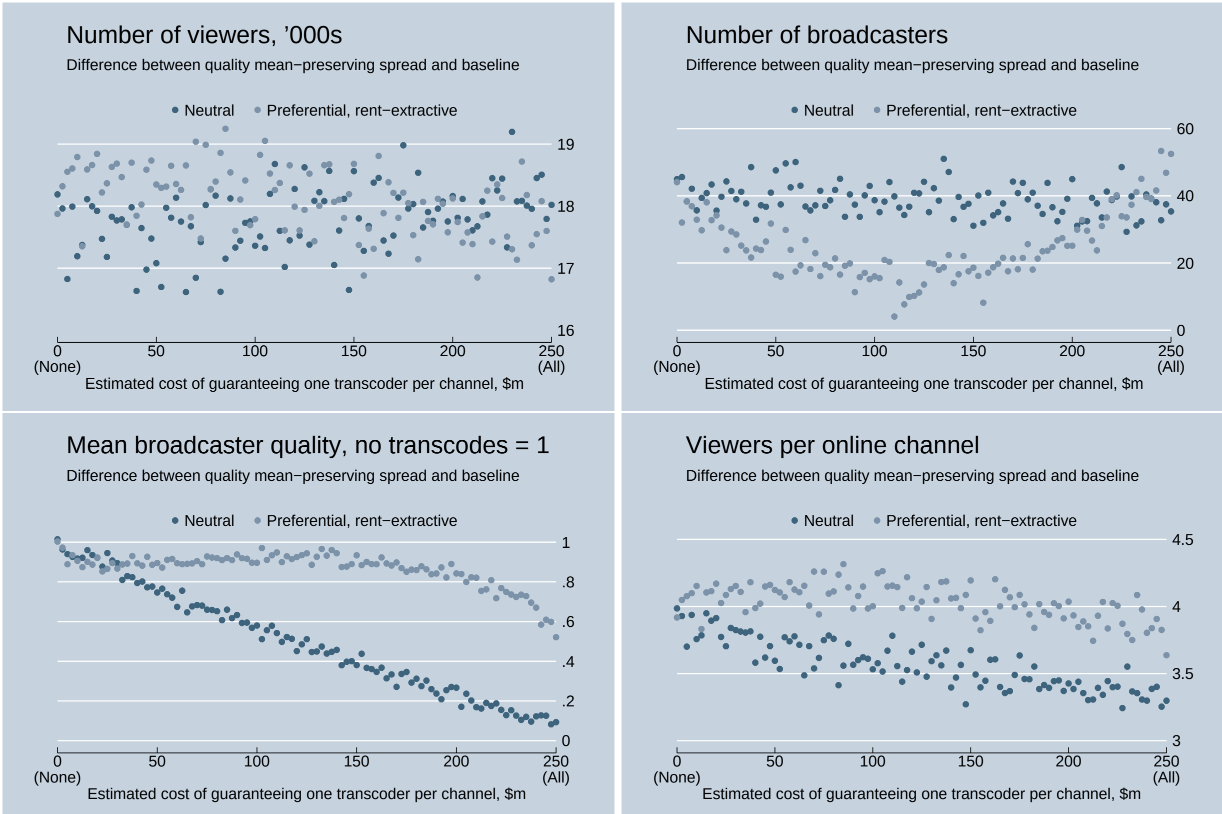
To explore the effects of heterogeneity, I consider a mean-preserving spread of quality. By adding a mean-zero shock, I increase the variance of the content provider's quality by 33%. Then, I compute the counterfactual equilibrium outcomes, and I compare the results with the outcomes of section 6.1. In this counterfactual comparison, the rent-extractive platform is maintained.

Figure 11 shows the results. Each dot represents the difference between (1) the equilibrium with a mean-preserving spread of quality, and (2) the equilibrium of section 6.1.

Increasing heterogeneity increases viewership under both neutral and preferential regimes. However, the preferential regime outperforms the neutral regime. Figure 11 shows the improvement is not because of more content providers, but because of the increase in the quality of online channels. Intuitively, when heterogeneity on the supply-side increases, the preferential regime allocates transcoders more efficiently, taking advantage of the fatter tail of the distribution. Therefore, if heterogeneity among content providers is pronounced, we would expect the benefits of prioritization to increase.

7 Concluding remarks

This paper asks if net neutrality is beneficial to consumers. The results suggest an immediate shift to neutrality could cost around 5% in consumer surplus, but the drop could



Note: Differences between averages of counterfactual simulations with and without a mean-preserving spread of quality. Averages computed from 5,000 equilibrium simulations. Figure 14 includes confidence intervals.

FIGURE 11: Effects of increasing quality heterogeneity

be as high as 10%, for plausible ranges of infrastructure levels. To hold consumer welfare constant and to be neutral, Twitch would need to significantly increase investment in transcoders.

In the worst-case scenario, the platform extracts surplus from content providers and leaves them indifferent between having preferential treatment and not having it. In this case, net neutrality does increase content provision. However, its effect on consumer surplus depends on the relative scarcity of transcoders. When this scarcity is pronounced, net neutrality decreases consumer welfare.

On the internet at large, the policymaker could observe that priority is, in fact, scarce. Only large platforms and internet service providers would have the infrastructure to support net neutrality at no cost to consumers. In general, however, the congestion generated by net neutrality likely outweighs the gains in content provision.

References

Becker, Gary S., Dennis W. Carlton, and Hal S. Sider. 2010. "Net neutrality and consumer welfare." *Journal of Competition Law and Economics*, 6(3): 497–519.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics*, 25(2): 242–262.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Caves, Kevin W. 2012. "Modeling the welfare effects of net neutrality regulation: A Comment on Economides and Tåg." *Information Economics and Policy*, 24(3-4): 288–292.

Choi, Jay Pil, and Byung-Cheol Kim. 2010. "Net neutrality and investment incentives." *The RAND Journal of Economics*, 41(3): 446–471.

Cisco. 2015. "The Zettabyte Era: Trends and Analysis."

Economides, Nicholas, and Benjamin E. Hermalin. 2012. "The Economics of Network Neutrality." *RAND Journal of Economics*, 43(4): 602–629.

Economides, Nicholas, and Joacim Tåg. 2012. "Network neutrality on the Internet: A two-sided market analysis." *Information Economics and Policy*, 24(2): 91–104.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance." *Annual Review of Economics*, 1(1): 255–294.

Greenstein, Shane, Martin Peitz, and Tommaso Valletti. 2016*a*. "Net Neutrality: A Fast Lane to Understanding the Trade-offs." *Journal of Economic Perspectives*, 30(2): 127–150.

Greenstein, Shane, Martin Peitz, and Tommaso Valletti. 2016*b*. "Net Neutrality Rules Will Make Winners and Losers Out of Businesses." *Harvard Business Review Digital Article*.

Guevara, Cristian Angelo, and Moshe E. Ben-Akiva. 2012. "Change of Scale and Forecasting with the Control-Function Method in Logit Models." *Transportation Science*, 46(3): 425–437.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47(1): 153–161.

Hortaçsu, Ali, and Joonhwi Joo. 2015. "Semiparametric Estimation of CES Demand System with Observed and Unobserved Product Characteristics."

Huynh-Thu, Q., and M. Ghanbari. 2008. "Scope of validity of PSNR in image/video quality assessment." *Electronics Letters*, 44(13): 800.

Lee, Robin S., and Tim Wu. 2009. "Subsidizing Creativity through Network Design: Zero-Pricing and Net Neutrality." *Journal of Economic Perspectives*, 23(3): 61–76.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics.* , ed. P. Zarembka, Chapter 4, 105–142. New York:Academic Press.

Newman, M. E. J. 2005. "Power laws, Pareto distributions and Zipf's law." *Power laws, Pareto distributions and Zipf's law. Contemporary physics*, 46(5): 323–351.

Nurski, Laura. 2012. "Net Neutrality, Foreclosure and the Fast Lane: An empirical study of the UK."

Pankert, Gregory, Andrea Faggiano, and Karim Taga. 2014. "The Future of the Internet: Innovation and Investment in IP Interconnection." Arthur D. Little.

Petrin, Amil, and Kenneth Train. 2010. "A Control Function Approach to Endogeneity in Consumer Choice Models." *Journal of Marketing Research*, XLVI(February): 3–13.

Pires, Karine, and Gwendal Simon. 2015. "YouTube Live and Twitch: A Tour of User-Generated Live Streaming Systems." 1–6. Portland, Oregon:ACM.

Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-Sided Markets: A Progress Report." *RAND Journal of Economics*, 37(3): 645–667.

Shivaldova, Veronika, Andreas Winkelbauer, and Christoph F. Mecklenbrauker. 2014. "Signal-to-noise ratio modeling for vehicle-to-infrastructure communications." *2014 IEEE 6th International Symposium on Wireless Vehicular Communications, WiVeC 2014 - Proceedings.*

Wu, Tim. 2003. "Network Neutrality, Broadband Discrimination." *Journal of Telecommunications and High Technology Law*, 2: 141.

A Demand-side first stages and robustness checks

Table 9 shows the first stages of demand estimations.

Additionally, Table 10 considers the following robustness checks: a selection correction based on the supply-side decision of Turning On, including controls; a Heckman (1979) selection correction based on the supply-side decision to Keep On (equation (3)); a selection correction for positive market shares, based on Hortaçsu and Joo (2015); and a nested logit model that allows for more flexible substitution patterns.

Column 3 of Table 10 considers the potential problem of zero market shares. Based on Hortaçsu and Joo (2015), I estimate a probit in a first stage, in which a positive market share is predicted by an English language dummy, a US country dummy, and the average video bitrate, as well as the covariates of the second stage. The results align with the selection corrections already performed.

In order to allow channels to be closer substitutes with each other than the outside option, the nested logit of column 4 considers the following model, in which the outside option is in one nest, and the rest of the channels are in a separate "inside" nest:

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = \mu_j + \alpha_1 \log \sigma_{jt} + \alpha_2 \tau_{jt} + \alpha_3 \log s_{jt|\mathrm{In}} + \xi_{jt}, \tag{6}$$

where $s_{jt|\mathrm{In}}$ is the market share of channel j as a fraction of the inside share (see Berry (1994)). By construction, this is an endogenous regressor. I instrument it with the percentage of channels in English at time t, and with this percentage interacted with an English language dummy. The English language is the most common language of the platform and commands a premium. All else equal, the demand for any single channel should be decreasing in the number of channels in English. Moreover, the percentage of other channels in English should not be relevant for the specific demand of channel j.

The results do not support a nested structure. As previously discussed, the classic intuition of price elasticities do not translate directly into noise elasticities. If channel j is noisy, it may be the case that consumers substitute directly into the outside option.

Table 11 considers a time trend in demand. The second column adds a relevant covariate when considering time-series variation. On February 14, 2014, a channel went unexpectedly viral and brought more viewers to the platform as well as salience in the media.[54] Column 2 also adds an instrument, which is an interaction of the Xbox One availability with the time trend. The first stages of these specifications are in Table 9, last columns.

Finally, Table 12 shows the first stages for Table 10.

B Supply-side first stages and robustness checks

Decision to Turn On. For completeness, Tables 13 presents the implied marginal effects from the preferred specification of Table 6. For reference, it also shows a linear probability model, as well as an additional specification that includes time fixed effects. Results are robust.

Decision to Keep On. Table 14 also shows the marginal effects of the preferred specification (column 4 of Table 7), omitting the coefficients of the control functions. For reference,

[54]As reported by `theguardian.com`, `bbc.com` and `businessinsider.com`.

it also includes a linear probability model. To be able to compare it with the conditional logit, the linear probability model includes the control functions as additional regressors. Columns 3 and 4 restrict the sample and perform a 2SLS estimation while column 5 emulates the conditional logit from column 4 of Table 7, for a restricted sample. The loss of power is evident. Pooling the sample helps to identify the coefficients of interest.

Table 15 shows the first stages that were used to construct the control function (based on equation (1′)), as well as the 2SLS of Table 14.

B.1 Control functions in the supply side

Consider a standard logistic error ε that can be decomposed into $\varepsilon \equiv \lambda\eta + \widetilde{\varepsilon}$, where $\lambda\eta$ is the control function (Petrin and Train, 2010). Consider the logit model $y = 1 \Leftrightarrow \theta x + \varepsilon$. Then $\theta x + \lambda\eta + \widetilde{\varepsilon}$ yields a scaled logit, since $\mathbb{V}\left[\widetilde{\varepsilon}\right] \leq \mathbb{V}\left[\varepsilon\right] = \pi^2/3$.

Normalize $\mathbb{V}\left[\eta\right] = 1$. Let $\widetilde{\lambda}$ be the coefficient of the control function in the scaled logit, and let $\widehat{\lambda}$ be the estimator of λ, the coefficient of the control function in the unscaled, original logit. Then,

$$\widetilde{\lambda} = \frac{\sqrt{\frac{\pi^2}{3}}}{\sqrt{\frac{\pi^2}{3} - \widehat{\lambda}^2}}\widehat{\lambda}^2 \Rightarrow \widehat{\lambda} = \sqrt{\frac{\frac{\pi^2}{3}\widetilde{\lambda}^2}{\frac{\pi^2}{3} + \widetilde{\lambda}^2}} \quad \therefore \frac{\sqrt{\frac{\pi^2}{3}}}{\sqrt{\frac{\pi^2}{3} - \widehat{\lambda}^2}} = \sqrt{1 + \frac{3\widetilde{\lambda}^2}{\pi^2}},$$

after simplifications. The last factor is the scaling factor that multiplies the coefficients of the scaled logit, in terms of the scaled coefficient $\widetilde{\lambda}$. To get the scaled coefficients in the original scale, we have to divide the scaled coefficients by the last factor. See also Guevara and Ben-Akiva (2012).

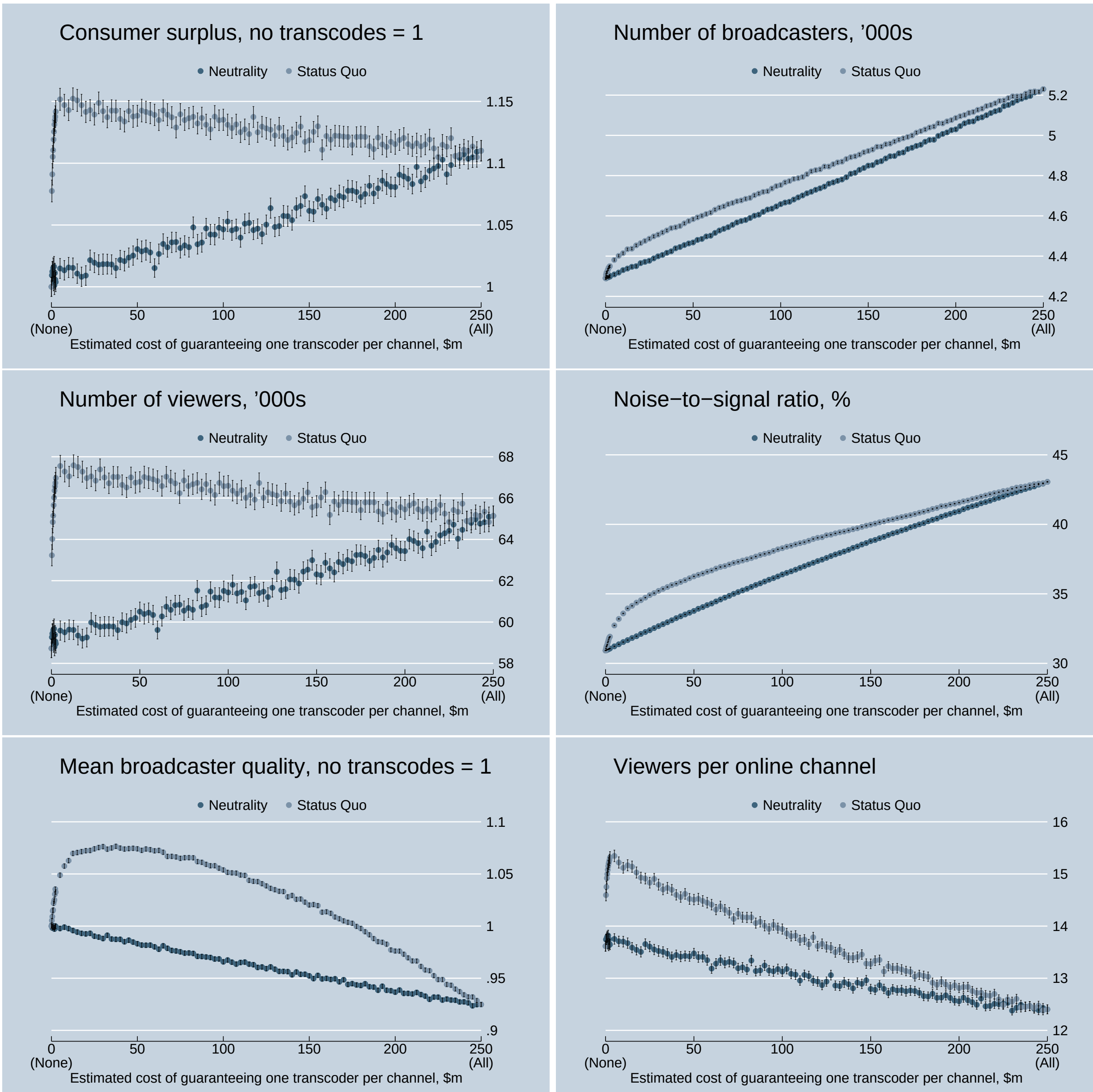
For the specific case of equation (3), $\lambda\eta \equiv \lambda_0\eta\mathbb{1}\{\text{Non-Partner}\}+\lambda_1\eta\mathbb{1}\{\text{Partner}\}$. Normalize, $\mathbb{V}\left[\eta\mathbb{1}\{\text{Non-Partner}\}\right] = \mathbb{V}\left[\eta\mathbb{1}\{\text{Partner}\}\right] = 1$. It follows that $\mathbb{V}\left[\lambda\eta\right] = \lambda^2 = \lambda_0^2 + \lambda_1^2$. This substitution yields the unscaled coefficients.

C Congestion first stages and robustness checks

Table 16 presents a selection correction based on Heckman (1979). Column 1 shows a 2SLS without fixed effects, and column 2 uses fixed effects in the second stage. To control for selection, both columns include an inverse Mills ratio. The probit of the first stage is in column 4 of Table 17. It is based on the decision to Keep On, equation (3). In this first stage probit, no fixed effects are included. Column 3 in Table 16 includes a time trend.

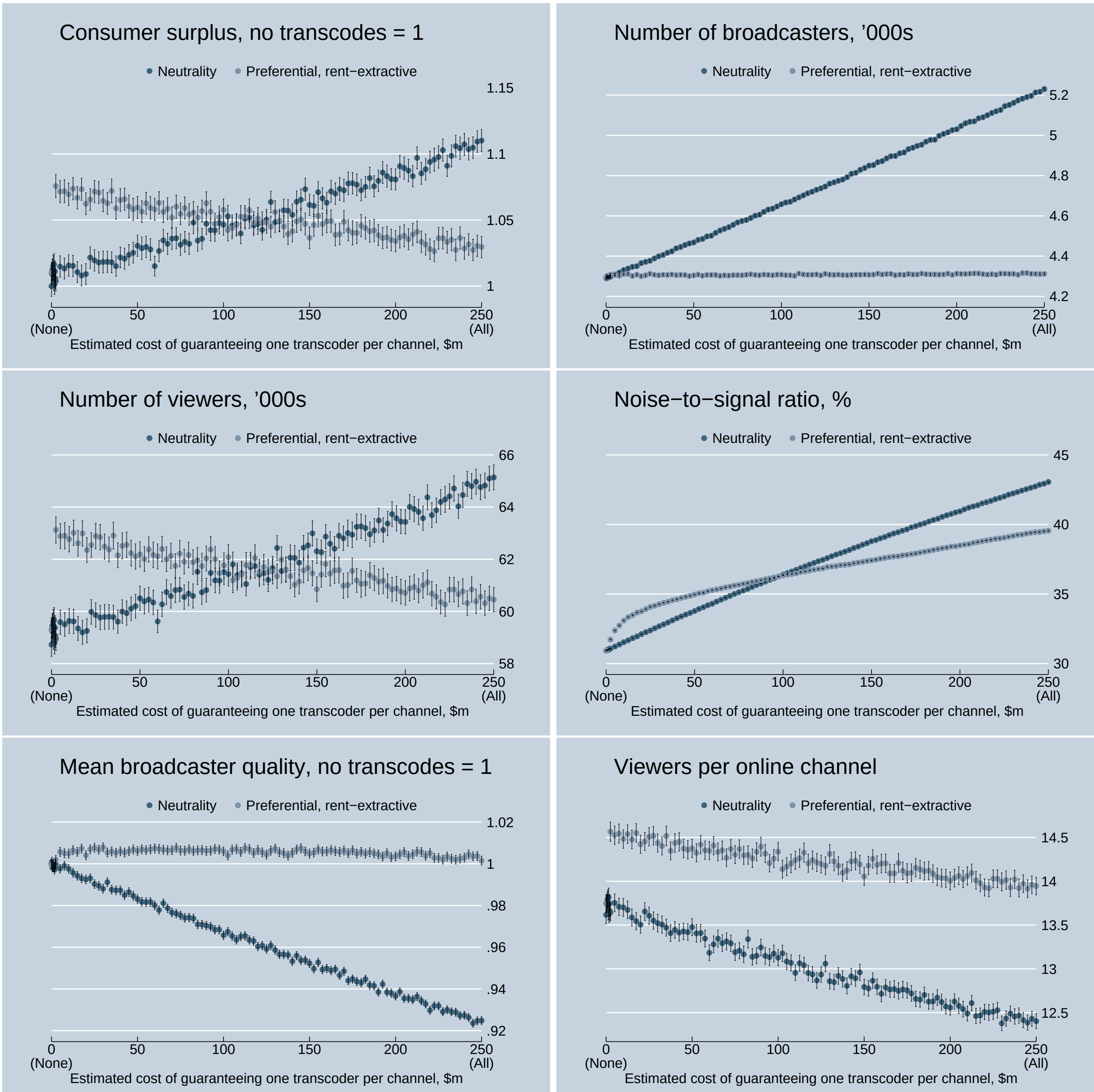
The PoP upgrades are overall negatively correlated with noise, and the Xbox One shock increases the aggregate, upload video bitrate.

D Additional counterfactual figures



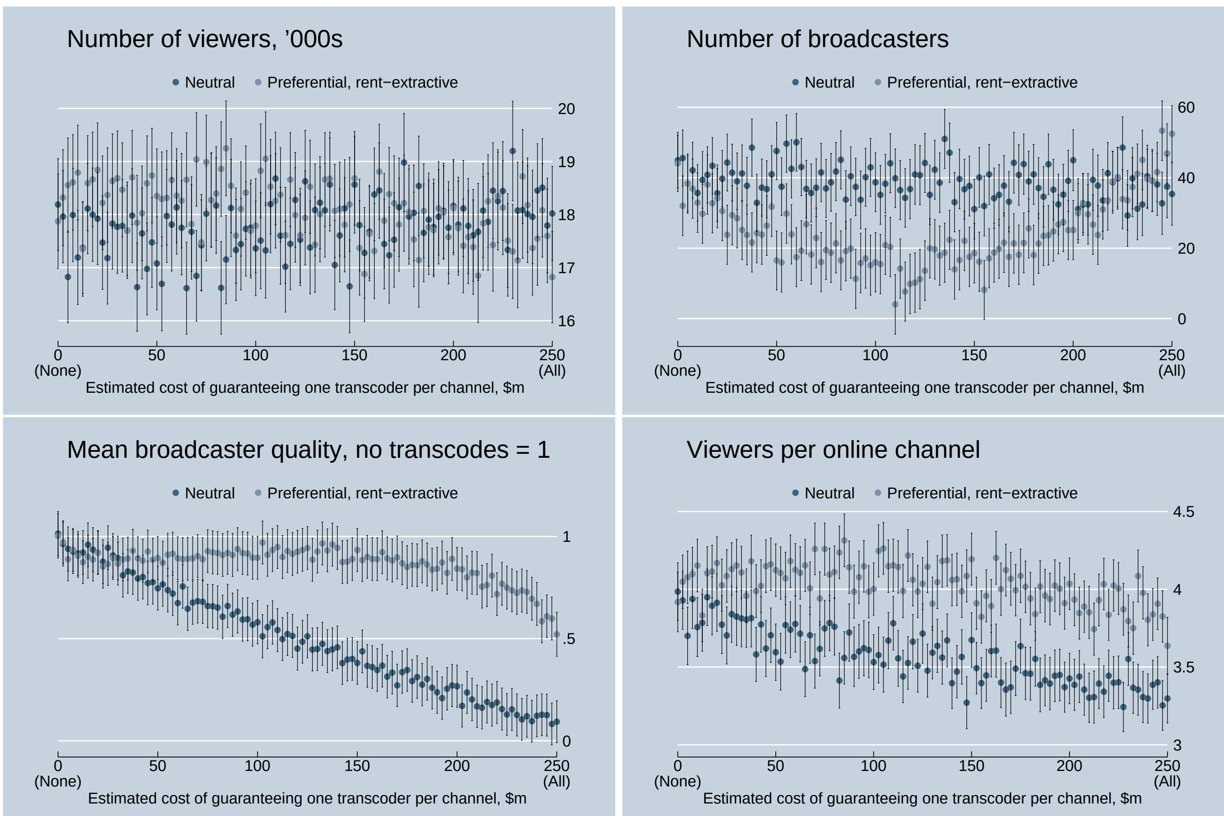
Note: Each dot represents the average outcome of 5,000 equilibrium simulations. Confidence intervals at 95% level. The red line marks the status quo, which is the preferential regime with enough transcoders for about 1% of the population.

FIGURE 12: Neutral v preferential regimes



Note: Each dot represents the average outcome of 5,000 equilibrium simulations. Confidence intervals at 95% level.

FIGURE 13: Neutral regime v preferential regime with rent-extractive platform



Note: Differences between averages of counterfactual simulations with and without a mean-preserving spread of quality. Averages computed from 5,000 equilibrium simulations. Confidence intervals at 95% level.

FIGURE 14: Effects of increasing quality heterogeneity

Table 1: Summary statistics

	Mean	Std	p10	p25	p50	p75	p90
All Sessions:							
Duration (min)	65	199	0	0	20	71	162
Mean Viewers	15.1	448.0	0.0	1.0	1.5	3.0	7.8
Max Viewers	22	705	0	1	2	4	11
Video Bitrate (kbps)	1498	1136	384	716	1157	2099	3036
Noise-to-Signal (%)	17	22	3	5	8	20	47
Observations	10,814,187						
Broadcasters	1,410,089						
Gaming:							
Duration (min)	57	113	0	0	20	70	157
Mean Viewers	15.0	459.5	0.0	1.0	1.5	2.9	7.0
Max Viewers	22	723	0	1	2	4	11
Video Bitrate (kbps)	1544	1135	436	769	1203	2125	3060
Noise-to-Signal (%)	17	22	3	5	8	20	46
Observations	10,252,065						
Broadcasters	1,360,177						
Non-Gaming:							
Duration (min)	202	712	0	0	25	113	427
Mean Viewers	17.2	103.6	0.0	0.1	1.5	6.7	40.5
Max Viewers	24	160	0	1	2	10	52
Video Bitrate (kbps)	651	756	67	193	441	877	1408
Noise-to-Signal (%)	22	24	3	7	14	29	50
Observations	562,122						
Broadcasters	50,414						
Weekend:							
Duration (min)	64	193	0	0	20	70	163
Mean Viewers	15.6	525.5	0.0	1.0	1.8	3.0	8.1
Max Viewers	23	841	0	1	2	4	12
Video Bitrate (kbps)	1489	1130	387	713	1151	2090	3024
Noise-to-Signal (%)	17	22	3	5	8	20	46
Observations	4,802,802						
Broadcasters	959,038						

Table 1: continued

	Mean	Std	p10	p25	p50	p75	p90
Non-Weekend:							
Duration (min)	65	203	0	0	21	71	161
Mean Viewers	14.6	374.8	0.0	1.0	1.3	2.8	7.5
Max Viewers	22	574	0	1	2	4	11
Video Bitrate (kbps)	1505	1141	383	719	1163	2107	3044
Noise-to-Signal (%)	17	22	3	5	8	21	47
Observations	6,011,385						
Broadcasters	1,037,408						
Partner:							
Duration (min)	199	237	10	60	152	274	420
Mean Viewers	804.2	2484.3	31.0	79.8	212.0	585.9	1559.9
Max Viewers	1174	3628	41	107	296	829	2259
Video Bitrate (kbps)	2340	1069	1023	1588	2273	3068	3564
Noise-to-Signal (%)	12	14	4	5	7	11	26
Observations	91,123						
Broadcasters	1,692						
Non-Partner:							
Duration (min)	63	198	0	0	20	70	159
Mean Viewers	8.4	380.3	0.0	1.0	1.5	2.9	7.0
Max Viewers	13	615	0	1	2	4	11
Video Bitrate (kbps)	1491	1134	382	713	1152	2093	3027
Noise-to-Signal (%)	17	22	3	5	8	21	47
Observations	10,723,064						
Broadcasters	1,409,047						

Notes: Statistics calculated for the complete sample. Data from Pires and Simon (2015), and from the author. Duration in minutes. Data was collected at 10 minute intervals, thus duration of "0" minutes implies duration of less than 10 minutes. Video bitrate in kbps. The noise-to-signal ratio is the video bitrate's coefficient of variation.

Table 2: Top countries, languages and timezones

Country	Freq	Language	Freq	Timezone	Freq
US	265,763	en	933,610	+01:00	311,803
GB	122,168	de	20,388	-08:00	209,975
DE	84,926	fr	13,585	+00:00	136,015
FR	52,476	ru	12,679	-05:00	109,203
SE	41,723	sv	8,098	-06:00	51,969
TW	40,076	zh-tw	7,622	+08:00	49,911

Note: The total number of channels is 1,045,214. There are 170 countries of origin in total, 43 total languages used in the platform, and 37 different timezones.

Table 3: Concurrent online channels, viewers, congestion, and duration, before and after technical change, with t-tests

	10 min windows				p-value:
	Before	After	Δ	S.E.	Ha: $\Delta \neq 0$
All channels	5,266	5,562	296	12.78	0.0000
Gaming	4,592	4,970	377	12.93	0.0000
Non-Gaming	674	593	-81	4.86	0.0000
Partner	136	164	28	0.77	0.0000
Non-Partner	5,130	5,398	268	12.80	0.0000
Viewers	308,524	312,364	3,840	2,230.72	0.0853
Gaming	284,550	290,583	6,033	2,219.56	0.0066
Non-Gaming	23,975	21,781	-2,194	81.18	0.0000
Partner	169,231	162,123	-7,107	949.61	0.0000
Non-Partner	139,294	150,241	10,947	1,854.13	0.0000
Bitrate/person	2,102	2,332	230	3.01	0.0000
Bitrate (upload)	1,533	1,629	96	2.02	0.0000
Gaming	1,663	1,741	79	1.41	0.0000
Non-Gaming	665	708	44	1.38	0.0000
Partner	2,379	2,465	86	1.69	0.0000
Non-Partner	1,511	1,604	93	1.96	0.0000
Duration (min)	66	64	-3	0.27	0.0000
Gaming	58	57	-1	0.17	0.0000
Non-Gaming	199	191	-7	2.63	0.0051
Partner	202	204	2	2.03	0.3412
Non-Partner	65	63	-3	0.27	0.0000
Noise-to-Signal (%)	19.8	20.1	0.29	0.02	0.0000
Gaming	19.5	19.9	0.39	0.02	0.0000
Non-Gaming	22.1	22.1	-0.02	0.03	0.5414
Partner	14.2	15.6	1.32	0.03	0.0000
Non-Partner	19.9	20.2	0.28	0.02	0.0000

Notes: S.E. refers to standard error of difference; p-value for two-sided t-test with unequal variances. Bitrates in kbps. Data aggregated in 10-minute period windows. Averages across 10-minute periods taken for before and after the change.

Table 4: Difference-in-Differences estimation

	Dependent variable	
	Broadcasters	Viewers
Gaming dummy	3,629	266,380
	(6)	(1,148)
Xbox dummy	-62	-1,778
	(4)	(54)
Interaction	372	7,486
	(14)	(2,202)
constant	483	15,692
	(2)	(23)

Notes: Bootstrap standard errors in parenthesis; 100 replications. All coefficients more than 99 percent significant. Treatment group is channels in gaming category; control is its complement. Xbox dummy indicates post March 11, 2014. Time series of 10-minute windows; 25,954 total observations.

Table 5: Demand estimations

Dep var: $\log(s_{jt}/s_{0t})$	Models (1)	(2)	(3)	(4)	(5)	(6)
$\log \sigma_{jt}$	-0.032^{***}	-0.013^{***}	-0.752^{***}	-0.449^{***}	-0.277^{***}	-0.210^{***}
	(0.006)	(0.001)	(0.221)	(0.064)	(0.068)	(0.023)
Transcoder	3.142^{***}	0.136^{***}	3.044^{***}	0.156^{***}	0.175^{***}	0.252^{***}
	(0.096)	(0.021)	(0.104)	(0.023)	(0.016)	(0.077)
Partner					−0.084	
					(0.064)	
Featured					2.342^{***}	
					(0.236)	
log Tenure					0.016^{**}	
					(0.007)	
Weekend					0.063^{***}	
					(0.002)	
After 6pm					0.060^{***}	
					(0.004)	
log Uptime					0.060^{***}	
					(0.001)	
cons	-13.830^{***}	-13.632^{***}	-15.581^{***}	-14.698^{***}	-15.037^{***}	-14.997^{***}
	(0.022)	(0.002)	(0.533)	(0.157)	(0.140)	(0.052)
Channel FE	No	Yes	No	Yes	Yes	Yes
2SLS	No	No	Yes	Yes	Yes	Yes
Restricted	.	.	.	.	.	Yes
Obs	12,065,648	12,065,648	12,065,648	12,065,648	10,405,683	1,372,811

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. Instruments for $\log \sigma_{jt}$: Xbox One broadcasting availability, PoP server upgrades. "Restricted" indicates that the sample was restricted to the first observation of a session; according to the supply side equation (2), channels do not take into account the potential effects of noise in the first observation, and, thus, selection does not occur. First stages reported in Table 9. Random sample of 20 percent of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 6: Decision to Turn On

	Models			
Dep var: $\text{Online}_{jt} = 1$	(1)	(2)	(3)	(4)
Partner	0.453^{***}	0.477^{***}	0.477^{***}	0.438^{***}
	(0.033)	(0.035)	(0.115)	(0.115)
Xbox, gaming	0.167^{***}	0.188^{***}	0.053^{***}	0.179^{***}
	(0.008)	(0.008)	(0.006)	(0.007)
log Downtime	-0.674^{***}	-0.668^{***}	-0.577^{***}	-0.550^{***}
	(0.002)	(0.002)	(0.001)	(0.001)
$\log\left(N_t^V/N_t^B\right)$		0.459^{***}	0.499^{***}	0.392^{***}
		(0.013)	(0.007)	(0.007)
Weekend				0.096^{***}
				(0.005)
6pm-1am				0.490^{***}
				(0.005)
log Tenure				-0.250^{***}
				(0.004)
cons	-1.639^{***}	-3.552^{***}		
	(0.012)	(0.056)		
Channel FE	No	No	Yes	Yes
R_p^2	0.11	0.11	0.08	0.08
Obs	198,158,676	198,158,676	179,454,520	179,454,520
Channels	30,595	30,595	27,497	27,497

Notes: Standard errors in parenthesis. Panel defined at channel id level. Time series of 10-minute windows. "Xbox, gaming" refers to the interaction of Xbox One availability and gaming dummies. The marginal effects and a linear probability model are shown in Table 13. Channels in the top 1% of most number of sessions are omitted for computational concerns. Logit with FE is a conditional logit in which fixed effects drop out. Periods with abnormally low participation, defined as in the lowest percentile, are omitted due to probably being caused by a site crash. Random sample of 4 percent of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 7: Decision to Keep On

Dep var: $\text{Online}_{j,t+1} = 1$	Models (1)	(2)	(3)	(4)
$\log n^{V}_{jt}$ (Non-Partner)	0.516^{***}	0.415^{***}	0.390^{***}	0.944^{***}
	(0.006)	(0.005)	(0.002)	(0.059)
$\log n^{V}_{jt}$ (Partner)	0.052^{***}	0.015	0.145^{***}	0.550^{***}
	(0.018)	(0.018)	(0.015)	(0.073)
Partner	1.568^{***}	1.421^{***}	1.203^{***}	1.963^{***}
	(0.106)	(0.100)	(0.090)	(0.211)
Xbox, gaming	-0.066^{***}	-0.048^{***}	-0.021^{***}	-0.042^{***}
	(0.009)	(0.006)	(0.004)	(0.004)
Weekend		-0.109^{***}	-0.093^{***}	-0.121^{***}
		(0.003)	(0.003)	(0.004)
After 6pm		-0.165^{***}	-0.185^{***}	-0.214^{***}
		(0.006)	(0.003)	(0.004)
log Uptime		0.082^{***}	-0.037^{***}	-0.070^{***}
		(0.001)	(0.000)	(0.004)
log Tenure		0.078^{***}	0.014^{***}	0.015^{***}
		(0.002)	(0.003)	(0.003)
Control F., non partner				-0.340^{***}
				(0.037)
Control F., partner				-0.038^{***}
				(0.008)
cons	1.297^{***}	0.591^{***}		
	(0.008)	(0.012)		
Channel FE	No	No	Yes	Yes
R^2_p	0.04	0.06	0.01	0.01
Obs	7,518,955	7,518,550	7,486,156	7,486,154
Channels	167,562	167,562	150,315	150,315

Notes: Panel defined at channel id level. Standard errors in parenthesis. Time series of 10-minute windows. All columns show logit models. The control function is the residual of a regression of $\log n^{V}_{jt}$ on the noise-to-signal ratio, indicators for partnerships and transcoders, fixed effects, and the rest covariates of the logit; the first stage is presented in Table 15. Channels in the top 1% of most number of sessions are omitted for computational concerns. Logit with FE is a conditional logit where fixed effects drop out. Random sample of 20% of data. The implied marginal effects and a Linear Probability Model are presented in Table 14. Standard errors not corrected for two-stage estimation. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 8: Noise estimations

Dep var: $\log \sigma_{jt}$	Models (1)	(2)	(3)	(4)
$\log B_t^{up}$	-0.098^{***}	0.832^{***}	0.327^{***}	0.438^{***}
	(0.009)	(0.093)	(0.061)	(0.081)
PoP1	-0.056^{***}	-0.075^{***}	-0.100^{***}	-0.116^{***}
	(0.004)	(0.004)	(0.003)	(0.002)
PoP2	0.001	-0.024^{***}	-0.017^{***}	-0.031^{***}
	(0.004)	(0.005)	(0.003)	(0.003)
PoP3	0.034^{***}	-0.019^{***}	-0.012^{***}	-0.019^{***}
	(0.004)	(0.005)	(0.003)	(0.005)
PoP4	-0.001	-0.071^{***}	-0.007	-0.040^{***}
	(0.004)	(0.008)	(0.005)	(0.006)
cons	-0.807^{***}	-15.588^{***}	-7.519^{***}	-9.073^{***}
	(0.140)	(1.480)	(0.964)	(1.289)
2SLS	No	Yes	Yes	Yes
Channel FE	No	No	Yes	Yes
Only first sess obs	.	.	.	Yes
R_a^2	0.00	.	0.00	0.00
Obs	68,744,096	68,744,096	68,744,096	8,639,245

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. B_t^{up} stands for aggregate upload video bitrate. The Xbox One shock is the excluded instrument. “Only first sess obs” refers to a restricted sample of first observations of broadcasting sesssions. Since noise is realized after turning on, the first few minutes should not be subject to selection. First stages shown in Table 17. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 9: Demand estimations, first stages A

	Models						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var:	$\log \sigma_{jt}$	$\log \sigma_{jt}$	$\log \sigma_{jt}$	$\log \sigma_{jt}$	$\log \sigma_{jt}$	$\log \sigma_{jt}$	$\log \sigma_{jt}$
Xbox Available	0.043***	0.014**	0.020***	0.028***	0.033***	0.053***	0.084
	(0.010)	(0.006)	(0.007)	(0.008)	(0.008)	(0.011)	(0.056)
PoP1	−0.069***	−0.094***	−0.086***	−0.113***	−0.096***	−0.040***	−0.021**
	(0.008)	(0.006)	(0.006)	(0.005)	(0.005)	(0.010)	(0.009)
PoP2	0.001	−0.007	−0.001	−0.020***	−0.008	0.026***	0.031***
	(0.009)	(0.006)	(0.006)	(0.005)	(0.005)	(0.009)	(0.009)
PoP3	−0.017*	−0.012**	−0.009	−0.021**	−0.020**	−0.013	−0.007
	(0.010)	(0.006)	(0.006)	(0.008)	(0.009)	(0.010)	(0.010)
PoP4	−0.017**	0.014***	0.022***	−0.012**	0.002	−0.003	0.018*
	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.009)	(0.010)
Transcoder	−0.137***	0.060***	0.049***	0.019	−0.054	−0.137***	−0.137***
	(0.045)	(0.014)	(0.013)	(0.069)	(0.176)	(0.045)	(0.045)
Partner			−0.038		0.031		
			(0.060)		(0.155)		
Featured			−0.102**		−0.033		
			(0.048)		(0.256)		
log Tenure			−0.048***		−0.050***		
			(0.005)		(0.003)		
Weekend			−0.003*		−0.008***		
			(0.002)		(0.002)		
After 6pm			0.012***		0.003		
			(0.002)		(0.003)		
log Uptime			−0.017***		−0.005***		
			(0.000)		(0.001)		
Trend						−0.001***	−0.002***
						(0.000)	(0.000)
Trend×Xbox							0.000
							(0.001)
TPP							0.043***
							(0.010)
cons	−2.360***	−2.341***	−1.981***	−2.102***	−1.882***	−2.352***	−2.342***
	(0.010)	(0.005)	(0.024)	(0.004)	(0.015)	(0.011)	(0.011)
Channel FE	No	Yes	Yes	Yes	Yes	No	No
Restricted	.	.	.	Yes	Yes	.	.
Obs	13,763,912	13,763,912	11,939,472	1,735,880	1,551,959	13,763,912	13,763,912

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. First 4 columns correspond to the first stages of Table 5. Column 5 adds controls to column 4 and corresponds to the first stage of column 1 in Table 10. "Restricted" indicates that the sample was restricted to the first observation of a session (see text). The last 2 columns correspond to the first stages of Table 11 in which further details are provided about its regressors. Random sample of 20% of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 10: Robustness checks for demand estimation, A

Dep var: $\log(s_{jt}/s_{0t})$	Models (1)	(2)	(3)	(4)	(5)
$\log \sigma_{jt}$	-0.125^{***}	-0.367^{***}	-0.312^{***}	-0.405^{***}	-0.383^{***}
	(0.028)	(0.045)	(0.065)	(0.068)	(0.064)
Transcoder	0.072	0.115^{***}	0.119^{***}	0.142^{***}	0.135^{***}
	(0.100)	(0.023)	(0.023)	(0.025)	(0.024)
Partner	0.048				
	(0.066)				
Featured	2.910^{***}				
	(0.279)				
log Tenure	-0.001				
	(0.003)				
Weekend	0.079^{***}				
	(0.001)				
After 6pm	0.073^{***}				
	(0.002)				
log Uptime	-0.001				
	(0.001)				
Heckman		-10.183^{***}			
		(0.051)			
Heckman 2			-0.429^{***}		
			(0.074)		
$\log s_{jt\|\text{In}}$				0.062	0.094
				(0.068)	(0.063)
cons	-14.932^{***}	-12.068^{***}	-14.268^{***}	-13.899^{***}	-13.494^{***}
	(0.054)	(0.113)	(0.172)	(0.847)	(0.775)
2SLS	Yes	Yes	Yes	Yes	Yes
Channel FE	Yes	Yes	Yes	Yes	Yes
Restricted	Yes	.	.	.	.
Obs	1,223,626	9,508,500	12,061,356	12,065,648	12,065,648

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. "Restricted" indicates that the sample was restricted to the first observation of a session (see text and Table 9 for the first stage). "Heckman" indicates that a control function was included as a regressor; based on Heckman (1979), I included the the inverse Mills ratio of a probit where the binary decision is to Keep On (see equation (3)). "Heckman 2" also indicates that a control function was included as a regressor; based on Hortaçsu and Joo (2015), I included the the inverse Mills ratio of a probit of an indicator for a channel's positive market share. This second probit includes video bitrate speeds, an english language dummy and a US country dummy as excluded variables. Both corrections solve a selection problem; the first is selection due to noise and the second is selection due to positive market shares. The final columns are a nested logit, where $s_{jt|\text{In}}$ is the market share of channel j as a fraction of the inside share, and is instrumented with the percentage of channels in English and with this percentage interacted with an English language dummy (see equation (6)). Probits and first stages are in Table 12. Random sample of 20% of data. Standard errors are not corrected for the two-stage estimation. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 11: Robustness checks for demand estimation, B

	Models	
Dep var: $\log(s_{jt}/s_{0t})$	(1)	(2)
$\log \sigma_{jt}$	-0.798^{***}	-0.611^{***}
	(0.243)	(0.219)
Transcoder	3.038^{***}	3.063^{***}
	(0.106)	(0.103)
Trend	0.000	-0.001^{***}
	(0.000)	(0.000)
TPP		0.078^{***}
		(0.020)
cons	-15.694^{***}	-15.234^{***}
	(0.580)	(0.522)
2SLS	Yes	Yes
Obs	12,065,648	12,065,648

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. "Trend" is a time trend, measured in days, and "TPP" is an indicator for dates post February, 14th 2014, when a channel unexpectedly went viral, as reported by `theguardian.com`, `bbc.com` and `businessinsider.com`. This channel increased both demand and the salience of the platform. In column 1, the noise-to-signal ratio, σ_{jt}, is instrumented with the Xbox One availability and PoP upgrades. Column 2 adds as an instrument an interaction between the time trend and the Xbox One shock. See Table 9 for the first stages. Random sample of 20% of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 12: Demand estimations, first stages B

	Models			
	(1)	(2)	(3)	(4)
Dep var:	$Online_{j,t+1} = 1$	$\mathbb{1}\{s_{jt} > 0\}$	$\log s_{jt\|\text{In}}$	$\log s_{jt\|\text{In}}$
$\log n^V_{jt}$ (Partner)	0.037***			
	(0.002)			
$\log n^V_{jt}$ (NonPartner)	0.263***			
	(0.000)			
Partner	0.725***	1.028***		
	(0.010)	(0.008)		
Weekend	−0.062***	0.214***		
	(0.001)	(0.001)		
Xbox Available	−0.018***	−0.007***	0.071***	0.070***
	(0.001)	(0.002)	(0.005)	(0.005)
Transcoder		0.360***	0.155***	0.155***
		(0.021)	(0.015)	(0.015)
Featured		0.490***		
		(0.005)		
PoP1		−0.004***	0.048***	0.048***
		(0.001)	(0.005)	(0.005)
PoP2		0.026***	−0.030***	−0.030***
		(0.001)	(0.005)	(0.005)
PoP3		−0.001	0.013***	0.014***
		(0.002)	(0.005)	(0.005)
PoP4		−0.011***	−0.005	−0.005
		(0.001)	(0.005)	(0.005)
$\log \overline{b}_j$ Non-transc		0.292***		
		(0.000)		
$\log \overline{b}_j$ Transc		0.264***		
		(0.003)		
English		−0.009***		
		(0.001)		
USA		0.160***		
		(0.001)		
Eng. Chann. %			−0.500***	
			(0.034)	
Eng. Chann. %, Eng.				−0.443***
				(0.035)
Eng. Chann. %, NonEng.				−0.766***
				(0.098)
cons	0.863***	−1.038***	−10.711***	−10.716***
	(0.001)	(0.003)	(0.029)	(0.028)
Model	Probit	Probit	2SLS	2SLS
Channel FE	No	No	Yes	Yes
Obs	24,171,486	27,507,970	24,171,486	24,171,486

Notes: Panels defined at channel id level. Standard errors in parenthesis, clustered at channel id where applicable. Time series of 10-minute windows. First stage probits and regressions of Table 10. Column 1 is a probit based on equation (3). Column 2 is a probit for an indicator of having a positive market share. Finally, $s_{jt|\text{In}}$ is the market share of channel j as a fraction of the inside share; $\overline{b}_j$ is the average video bitrate of channel j, by transcoded status; the percentage of channels in English is a contemporaneous one and, in the last column, it is interacted with a dummy for English language of channel j. Random sample of 20 percent of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 13: Decision to Turn On: Robustness checks

	Models		
Dep var: $\text{Online}_{jt} = 1$	(1)	(2)	(3)
Partner	0.0003***	0.0010***	0.0004*
	(0.0001)	(0.0003)	(0.0002)
Xbox, gaming	0.0001***	0.0002***	
	(0.0000)	(0.0000)	
log Downtime	−0.0004***	−0.0010***	−0.0012***
	(0.0000)	(0.0000)	(0.0000)
$\log\left(N_t^V/N_t^B\right)$	0.0003***	0.0003***	
	(0.0000)	(0.0000)	
Weekend	0.0001***	0.0002***	
	(0.0000)	(0.0000)	
6pm-1am	0.0004***	0.0006***	
	(0.0000)	(0.0000)	
log Tenure	−0.0002***	−0.0001***	−0.0001***
	(0.0000)	(0.0000)	(0.0000)
cons		0.0092***	
		(0.0001)	
Channel FE	Yes	Yes	Yes + time FE
Model	margins	LPM	LPM
R_a^2	.	0.00	0.00
Obs	179,454,520	198,158,676	198,158,676
Channels	27,497	30,595	30,595

Notes: Standard errors in parenthesis. Panel defined at channel id level. Time series of 10-minute windows. "Xbox, gaming" refers to the interaction of Xbox One availability and gaming dummies. Channels in the top 1% of most number of sessions are omitted for computational concerns. Logit with FE is a conditional logit in which fixed effects drop out. Periods with abnormally low participation, defined as in the lowest percentile, are omitted due to probably being caused by a site crash. First column shows the marginal effects of the preferred specification (last column of Table 6). The second column is a Linear Probability Model, included for reference. The third column is also a LPM with time fixed effects and channel fixed effects. Random sample of 4% of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 14: Decision to Keep On: Robustness checks

Dep var: $\text{Online}_{j,t+1} = 1$	(1)	(2)	Models (3)	(4)	(5)
$\log n_{jt}^V$ (Non-Partner)	0.161***	0.046***	0.049**		
	(0.010)	(0.017)	(0.020)		
$\log n_{jt}^V$ (Partner)	0.094***	0.033*		−0.010	−0.793*
	(0.012)	(0.018)		(0.015)	(0.450)
Partner	0.335***	0.064***		−0.001	−0.063
	(0.036)	(0.016)		(0.003)	(0.047)
Xbox, gaming	−0.007***	−0.002***	−0.003**	−0.001	0.068**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.032)
Weekend	−0.021***	−0.009***	−0.009***	0.004***	0.172***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.026)
After 6pm	−0.036***	−0.018***	−0.019***	−0.003	−0.032
	(0.001)	(0.001)	(0.001)	(0.003)	(0.040)
log Uptime	−0.012***	−0.005***	−0.005***	−0.006*	−0.372***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.058)
log Tenure	0.002***	0.002***	0.002***	0.004	−0.002
	(0.001)	(0.001)	(0.001)	(0.014)	(0.147)
Control F., non partner		−0.006			
		(0.011)			
Control F., partner		−0.002		−0.038***	0.749***
		(0.002)		(0.008)	(0.261)
cons		0.841***	0.840***	1.036***	
		(0.015)	(0.015)	(0.100)	
Channel FE	Yes	Yes	Yes	Yes	Yes
Desc.	margins	LPM	LPM	LPM	Logit
2SLS	.	.	Yes	Yes	.
Restricted	.	.	NonPartner	Partner	Partner
$R^2_{p,a}$	0.00	0.02	0.02	0.00	0.04
Obs	7,486,154	7,518,548	7,330,454	188,094	188,094
Channels	150,315	167,562	167,322	240	240

Notes: Panel defined at channel id level. Standard errors in parenthesis. Time series of 10-minute windows. All columns show logit models. The control function is the residual of a regression of $\log n_{jt}^V$ on the noise-to-signal ratio, indicators for partnerships and transcoders, fixed effects, and the rest covariates of the logit; the first stage is presented in Table 15. Channels in the top 1% of most number of sessions are omitted for computational concerns. Logit with FE is a logit conditional on one success. "Margins" indicates that the column shows the marginal effects of the preferred specification of Table 7. "LPM" indicates a Linear Probability Model, included for reference. Also shown in Table 15 are the first stages of the 2SLS, which also instruments $\log n_{jt}^V$ with the log noise-to-signal ratio, and indicators for transcoders and features. Random sample of 20% of data. Standard errors not corrected for two-stage estimation where control functions are used. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 15: Decision to Keep On: First stages

Dep var: $\log n_{jt}^{V}$	Models (1)	(2)	(3)
$\log \sigma_{jt}$	-0.006^{***}	-0.006^{***}	-0.008^{*}
	(0.001)	(0.001)	(0.005)
Transcoder	0.172^{***}	0.163^{***}	0.340^{***}
	(0.015)	(0.015)	(0.093)
Featured	2.360^{***}	2.370^{***}	
	(0.238)	(0.238)	
Partner	-0.109^{**}		-0.162^{**}
	(0.053)		(0.072)
Xbox, gaming	0.039^{***}	0.044^{***}	-0.073^{**}
	(0.007)	(0.007)	(0.034)
Weekend	0.052^{***}	0.052^{***}	0.030
	(0.002)	(0.003)	(0.019)
After 6pm	0.058^{***}	0.058^{***}	0.049^{*}
	(0.004)	(0.004)	(0.027)
log Uptime	0.061^{***}	0.059^{***}	0.199^{***}
	(0.000)	(0.000)	(0.007)
log Tenure	−0.001	−0.002	0.089
	(0.005)	(0.005)	(0.241)
cons	0.786^{***}	0.731^{***}	3.185^{**}
	(0.028)	(0.027)	(1.562)
Channel FE	Yes	Yes	Yes
Partners only	.	NonPartner	Partner
R_a^2	0.07	0.07	0.26
Obs	7,518,548	7,330,454	188,094
Channels	167,562	167,322	240

Notes: Panel defined at channel id level. Standard errors in parenthesis. Time series of 10-minute windows. Channels in the top 1% of most number of sessions are omitted for computational concerns. The residuals of column 1 were used to construct control functions which are included in the models of Tables 7 and 14. Columns 2 and 3 show the first stages of the 2SLS used in Table 14. Random sample of 20% of data. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 16: Noise equation estimations: Robustness checks

	Models		
Dep var: $\log \sigma_{jt}$	(1)	(2)	(3)
$\log B_t^{up}$	0.512***	0.252***	1.994***
	(0.105)	(0.069)	(0.171)
PoP1	−0.071***	−0.102***	−0.013***
	(0.005)	(0.003)	(0.005)
PoP2	−0.017***	−0.014***	0.050***
	(0.006)	(0.004)	(0.003)
PoP3	−0.013**	−0.011***	−0.009**
	(0.006)	(0.004)	(0.004)
PoP4	−0.034***	0.013**	−0.047***
	(0.009)	(0.006)	(0.007)
Heckman	0.225***	0.330***	
	(0.045)	(0.011)	
Trend			−0.007***
			(0.000)
cons	−10.585***	−6.449***	−33.744***
	(1.671)	(1.092)	(2.717)
2SLS	Yes	Yes	Yes
Channel FE	No	Yes	Yes
Only first sess obs	.	.	Yes
R_a^2	.	0.00	0.00
Obs	48,314,880	48,314,880	8,639,245

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. B_t^{up} stands for aggregate upload video bitrate. "Heckman" indicates a Heckman (1979) two-stage correction. Its first stage is a probit based on the decision to Keep On in equation (3). I then included the predicted inverse Mills ratio in the second stage. The first stage probit is shown in Table 17 as well as the first stages of the 2SLS models. sesssions. Since noise is realized after turning on, the first few minutes should not be subject to selection. Standard errors not corrected for two-stage estimation in first two columns. Stars: *** significant at the 1% level; ** at 5%; * at 10%.

Table 17: Noise estimations, First stages

	Models						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep var:	$\log B_t^{up}$	$\log B_t^{up}$	$\log B_t^{up}$	Keep On	$\log B_t^{up}$	$\log B_t^{up}$	$\log B_t^{up}$
Xbox Available	0.050***	0.047***	0.043***	−0.040***	0.051***	0.048***	0.034***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
PoP1	0.021***	0.019***	0.015***	−0.019***	0.020***	0.019***	−0.012***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
PoP2	0.027***	0.027***	0.027***	0.002**	0.028***	0.028***	0.004***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
PoP3	0.012***	0.013***	0.016***	−0.007***	0.012***	0.013***	0.011***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
PoP4	0.069***	0.066***	0.068***	0.082***	0.070***	0.067***	0.053***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Heckman					0.055***	0.060***	
					(0.001)	(0.002)	
$\log n_{jt}^V$ (Partner)				0.035***			
				(0.001)			
$\log n_{jt}^V$ (Non-Partner)				0.266***			
				(0.000)			
Partner				0.780***			
				(0.006)			
Weekend				−0.069***			
				(0.000)			
Transcoder				−0.041***			
				(0.002)			
Featured				0.008***			
				(0.002)			
Trend							0.001***
							(0.000)
cons	15.891***	15.893***	15.902***	0.878***	15.883***	15.883***	15.892***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Channel FE	No	Yes	Yes	No	No	Yes	Yes
Model	OLS	OLS	OLS	Probit	OLS	OLS	OLS
Only first sess obs	.	.	Yes	.	.	.	Yes
$R^2_{a,p}$	0.19	0.13	0.14	0.05	0.21	0.14	0.14
Obs	68,981,883	68,981,883	8,876,901	60,574,441	48,314,974	48,314,974	8,876,901

Notes: Standard errors in parenthesis, clustered at channel id. Time series of 10-minute windows. B_t^{up} stands for aggregate, upload video bitrate. Column 4 is a first stage probit, used to construct an inverse Mills ratio to control for selection (see Table 16). "Heckman" indicates the inclusion of said inverse Mills ratio in the regression. "Only first sess obs" refers to a restricted sample of first observations of broadcasting sesssions. Stars: *** significant at the 1% level; ** at 5%; * at 10%.