

# The Tragedy of the Last Mile: Congestion Externalities in Broadband Networks\*

Jacob B. Malone<sup>†</sup>    Aviv Nevo<sup>‡</sup>    Jonathan W. Williams<sup>§</sup>

PRELIMINARY AND INCOMPLETE: April 2016

## Abstract

We flexibly estimate demand for residential broadband accounting for congestion externalities that arise among consumers due to limited network capacity and dynamics in usage resulting from nonlinear pricing. To estimate demand, we build a dynamic model of consumer choice and exploit exogenous variation in the timing of network upgrades to capacity. Our high frequency data permits insight into temporal patterns in usage across the day that are impacted by network congestion, and how usage responds to efforts to mitigate congestion. We show that usage is highly responsive to reductions in congestion, and there is substantial heterogeneity in the response across consumers. Using the model estimates, we then calculate the welfare loss to consumers associated with the existing externalities and compare it to the cost of eliminating them.

**Keywords:**

**JEL Codes:** L11, L13, L96.

---

\*We are grateful to the North American Internet Service Providers that provided the data used in this paper. We thank participants in several seminars, Ron Reuss, Sanjay Patel, and ..... for insightful comments. Jim Metcalf and Craig Westwood provided expert IT support for this project. We gratefully acknowledge support from CableLabs and NSF (grants SES-1324851 and SES-1324717).

<sup>†</sup>Department of Economics, University of Georgia, jbmalone@uga.edu.

<sup>‡</sup>Department of Economics, Northwestern University, nevo@northwestern.edu.

<sup>§</sup>Department of Economics, University of North Carolina at Chapel Hill, jonwms@unc.edu.

# 1 Introduction

The Internet is now an ever-present part of society, and the demand for online content, especially over-the-top (OTT) video, is soaring. Internet Service Providers (ISPs) choose to invest and meet this demand when there is incentive to do so. An industry estimate places private broadband investment around \$1.3 trillion between 1996 and 2013, or about \$75 billion per year.<sup>1</sup> Historically broadband investment has been financed by private firms, but its importance is now leading some local governments to pursue municipal broadband and other public funding to support further investment and competition.

In this paper, we provide a key input to understanding how responsive consumers are to network congestion by estimating demand using a novel data set and variation in network congestion and prices. Congested areas of the network are prime candidates for investment because a consistently poor network could lead to welfare losses for consumers, ISPs, and third-parties. In particular, we focus on congestion abatement and its value to consumers. We believe these results are of particular importance to any public policy debate that evaluates the value created by broadband investment. For example, as a part of the Charter/Time Warner Cable merger review, broadband investment to modernize and expand the network is a likely condition for approval.<sup>2</sup>

The unique data at the center of this work are made available by a North American ISP. These data include hourly observations of Internet usage and network conditions for roughly 45,000 subscribers from February 2015 through December 2015. At the daily level, we are able to uniquely map an account to a cable modem and active Internet plan. For each Internet plan, we observe the price, advertised speeds<sup>3</sup> (downstream and upstream), usage allowance, and overage fees. All data tiers charge for data overages at the same per GB rate. The average subscriber in our data uses 2.3 gigabytes (GB) per day, pays \$58.89 for a 22 megabit per second (Mbps) downstream connection, and a 267 GB monthly usage allowance.

Network congestion occurs when demand pushes or exceeds the network’s limitations – similar to how video streaming performance degrades if too many people utilize the same WiFi connection. Aside from downstream and upstream speeds, the Federal Communications Commission (FCC) recognizes *latency* (how long it takes requests to move across the Internet) and *packet loss* (roughly, the percentage of requests that fail to make it to their destination) as two important metrics of network performance.<sup>4</sup> Buffering video

---

<sup>1</sup>See USTelecom’s estimates at <http://www.ustelecom.org/broadband-industry-stats/investment/historical-broadband-provider-capex> and page 15 of the FCC’s 2015 Broadband Progress Report found at <https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2015-broadband-progress-report>.

<sup>2</sup>See the State of New York Public Service Commission’s Case 15-M-0388 on “Joint Petition of Charter Communications and Time Warner Cable for Approval”, released on January 8, 2016.

<sup>3</sup>Speed is measured in Megabits per second (Mbps). For reference, a 20 Mbps downstream connection would download a 4 gigabyte file, or roughly one high-definition movie, in about 27 minutes.

<sup>4</sup>See the FCC’s 2015 Measuring Broadband Report at <https://www.fcc.gov/reports-research/>

streams, websites failing to load, and being disconnected from an online video game are common examples of how congestion might affect a consumer. Moreover, due to differences in implementation, such activities as sending email are more resilient to congestion than others like video streaming.

ISPs can invest in the physical network in two primary ways.<sup>5</sup> First, network investment can expand the current network, usually in rural and poorer communities, where costs can be greater and broadband is rarer. The government actively promotes this type of investment. Some communities such as Chattanooga, TN and Lafayette, LA have voted for municipal broadband, arguing it promotes competition and investment.<sup>6</sup> Additionally, many politicians have plans similar to President Obama’s ConnectHome<sup>7</sup> that address how the government will support network investment in these areas.

Second, an ISP can invest in the existing network by increasing capacity and speeds. Here, existing customers are receiving a better experience from the improved network quality. Network performance is of importance to the FCC, which it tracks in its annual Measuring Broadband America reports.<sup>8</sup> In these reports, various network test results of several popular US ISPs are released to promote transparency in the quality and options of broadband available to subscribers. This type of network improvement is what typically abates congestion and is most relevant to this research.

A New York Times article<sup>9</sup> describes network investment as having two types of costs: the cost of connecting people’s houses, and the cost of delivering bandwidth to these networks. In general, updating the links between people is the more costly of the two, since a *node*<sup>10</sup> only provides a fixed amount of shared bandwidth to subscribers. Estimating the cost and structure of bandwidth prices are more complicated since they depend on an ISP’s ownership of infrastructure, peering, and interconnection agreements.

*Node splits* are a common way ISPs invest in the core network to improve capacity and lower congestion for a group of subscribers. A node is a common place for bottlenecks to occur and are what commonly demarcate local, “last mile” networks. When a node is split, its subscribers are distributed evenly across two new nodes, where network

---

reports/measuring-broadband-america/measuring-broadband-america-2015.

<sup>5</sup>A third way an ISP could invest in the network is by improving “upstream” relations through various peering and interconnection agreements. Typically, these agreements result in faster, less-congested routes between certain destinations – for example, an ISP and Netflix or Hulu.

<sup>6</sup>See the White House release “Community-Based Broadband Solutions ” at [https://www.whitehouse.gov/sites/default/files/docs/community-based\\_broadband\\_report\\_by\\_executive\\_office\\_of\\_the\\_president.pdf](https://www.whitehouse.gov/sites/default/files/docs/community-based_broadband_report_by_executive_office_of_the_president.pdf).

<sup>7</sup>The program’s website can be found at <http://connecthome.hud.gov> and a White House release summarizing the program is found at <https://www.whitehouse.gov/the-press-office/2015/07/15/fact-sheet-connecthome-coming-together-ensure-digital-opportunity-all>.

<sup>8</sup>The latest Measuring Broadband America report can be found at <https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-2015>.

<sup>9</sup>See “The Cost of Downloading All Those Videos” by Saul Hansell, released April 20, 2009, at [http://bits.blogs.nytimes.com/2009/04/20/the-cost-of-downloading-all-those-videos/?\\_r=0](http://bits.blogs.nytimes.com/2009/04/20/the-cost-of-downloading-all-those-videos/?_r=0).

<sup>10</sup>A *node* is a network device that connects a group of subscribers to the rest of the operator’s network.

conditions should be improved. Many operators target nodes to be split once average utilization exceeds certain thresholds. We observe 5 node splits in our data and use these events to compare before-and-after congestion and subscriber usage. After a split, average daily usage increases by 7% and packet loss, our measure of congestion, drops by 27%. This suggests there is value to consumers from a less congested network.

The value to consumers of less congestion only captures part of the rent created by the investment: some goes back to the ISP and the rest goes to third-parties. In fact, since the ISP is unable to fully capture these rents, private investment is marginally discouraged. Moreover, recent Title II and net neutrality regulation by the FCC has created uncertainty on the future of the industry, which could depress future investment. Tom Wheeler, the current FCC chairman, declares Title II will have no effect on investment, while other commissioners are doubtful.<sup>11</sup>

Our model of subscriber Internet consumption builds on the framework of Nevo et al. (2016) with the notable difference being the inclusion of network congestion and its impact on plan choice and consumption. Similarly, our estimation relies on variation in prices and speeds across plans and (shadow) price variation across the billing cycle that is created by usage-based pricing. We also utilize variation in a subscriber’s observed packet loss to estimate the effect of congestion.

The price variation arising from usage-based pricing is a result of its three-part tariff structure: a subscriber pays a fixed fee each month, and if the associated usage allowance is exceeded, she is charged at a per GB rate thereafter. While overage fees are only assessed if a subscriber exceeds the usage allowance, a forward-looking subscriber understands today’s consumption marginally increases the likelihood of exceeding the usage allowance before the end of the billing cycle – this is a function of how many days remain in the billing cycle and what fraction of the usage allowance has been used previously in the cycle. We incorporate these dynamics in our model similar to Nevo et al. (2016) by allowing consumers to make daily consumption decisions across a billing cycle.

We also use variation in network congestion to identify a subscriber’s sensitivity to poor network states. Our hourly data contain *packet loss*, or the percentage of total packets requested that are either dropped or delayed, at the subscriber level, which we use to proxy for network congestion. As mentioned previously, packet loss is one statistic the FCC uses to benchmark network performance across ISPs.

There are four main advantages to using packet loss over other variables such as a node’s utilization to measure congestion in our model. First, we observe packet loss at a subscriber level, so it is not an aggregate network statistic. Second, we observe wide cross-sectional variation in packet loss across subscribers. Third, packet loss is positively

---

<sup>11</sup>See Chairman Wheeler’s article “This is How We Will Ensure Net Neutrality” at <http://www.wired.com/2015/02/fcc-chairman-wheeler-net-neutrality/>, for example, and Commissioner Pai’s remarks on “Declining Broadband Investment” at <https://www.fcc.gov/document/comm-pai-remarks-declining-broadband-investment>.

correlated with other common congestion variables. Fourth, our ISP invested in its existing network throughout the year, so we are able to exploit time series variation in our panel as well. Our ISP’s network would rate as the third worst in average subscriber packet loss in the latest Measuring Broadband America. Coupled with the aforementioned core network investments by the operator over 2015, this sample offers a wide range of variation in network conditions, ideal for this analysis.

We estimate this finite horizon, dynamic choice model by solving the dynamic problem once for a large number of types. The solution to these dynamic problems is then used to estimate the distribution of types over our sample by minimizing the error between observed and optimal behavior across types. In general, the estimated marginal and joint distributions illustrate the strength of the flexibility built into our estimation approach. Compared to Nevo et al. (2016)’s concentrated type distribution, ours is much more uniform.

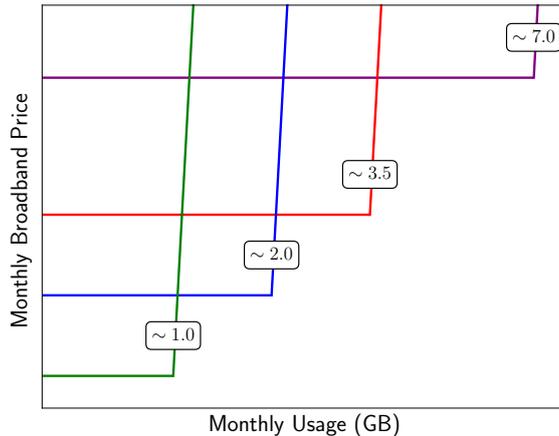
These demand estimates are used to measure the value to subscribers when network congestion is eliminated entirely. This is the case where a subscriber’s provisioned speed is always realized. We find the improved network conditions encourage some subscribers to downgrade to cheaper plans, but the loss in revenue from this is entirely offset by an increase in consumer surplus. Subscribers’ realized speeds increased by roughly 19% with each additional Mbps of speed being valued at roughly \$2.87. These results suggest that when public policy focuses on network investment, congestion abatement should be considered because of this positive value enjoyed by subscribers.

This paper is most closely related to a literature that studies the demand of residential broadband. Recent examples are Nevo et al. (2016), Malone et al. (2016), and Malone et al. (2014) that use similar high-frequency data to study subscriber behavior. However, this literature dates back to the early 2000s with Varian (2002) and Edell and Varaiya (2002), who run experiments where consumers face different prices for varying allowances and speeds. Goolsbee and Klenow (2006) estimate the benefit to residential broadband; Hitte and Tambe (2007) show Internet usage increases by roughly 22 hours per month when broadband is introduced. Other related papers are Lambrecht et al. (2007), Dutz et al. (2009), Rosston et al. (2013), and Greenstein and McDevitt (2011).

## 2 Data

The data for our analysis come from a representative sample of 46,667 North American broadband subscribers. The metropolitan area where the subscribers are drawn from have demographic characteristics that are similar to the entire US population, its average income is within 10% of the national average and the demographic composition is similar to the overall US population. Therefore, we expect the insights from our analysis to have external validity in other North American markets. The data include hourly subscriber usage and details of network conditions for February–December 2015.

Figure 1: Internet Plan Features



*Note:* This figure represents the relationship between monthly usage and price for the ISP’s four Internet plans. Since this ISP has implemented usage-based pricing, there is a set usage allowance for each plan. Once this usage allowance is exceeded, the subscriber is billed on a per GB basis. The overage rate is the same across all four plans. The label that intersects each plan’s line represents the relative differences in speeds.

Our data set is constructed from three primary sources. The first source is Internet Protocol Detail Records (IPDR), which report hourly counts of downstream and upstream bytes, packets passed, and packets dropped/delayed by each cable modem.<sup>12</sup> IPDR also record a cable modem’s *node*, a device that connects a set of customers to the rest of the operator’s network. The second data source is average hourly utilization by node. The last are billing records by customer, where service plan details (e.g., speed, usage allowance, prices) are included. These data sets are linked by a customer’s anonymized account number, which maps uniquely to a cable modem by day. Using an account number as our unique identifier allows us to follow customers across hardware changes within the sample.

## 2.1 Sample, Internet Plans, and Subscriber Usage

Our panel starts on February 1, 2015 and ends on December 31, 2015 and includes 309,307,896 *subscriber-day-hour* observations. For each observation, we observe downstream/upstream bytes and the total number of packets passed and dropped/delayed. At a daily frequency, we observe each subscriber’s plan and the mapping of consumers to nodes in the network.

The ISP sells Internet access via a menu of plans with more expensive plans including both faster access speeds and larger usage allowances. Overages are charged on a per GB basis after the usage allowance is exceeded. The relationship between monthly usage (GB) and monthly price (\$) across the plans is shown in Figure 1. The average subscriber pays

<sup>12</sup>All cable modem hardware identifiers are hashed to preserve anonymity.

Table 1: *Daily Usage Distributions by Internet Plan Tier*

	<i>Tier 1</i>	<i>Tier 2</i>	<i>Tier 3</i>	<i>Tier 4</i>	<i>All</i>
Mean	1.4 GB	3.4 GB	5.4 GB	8.2 GB	2.3 GB
Std. Dev.	2.9	5.0	7.3	10.4	4.5
25 <sup>th</sup> %tile	0.0	0.3	0.6	1.3	0.1
Median	0.4	1.5	3.1	5.3	0.6
75 <sup>th</sup> %tile	1.5	4.7	7.6	11.4	2.7
90 <sup>th</sup> %tile	4.1	9.0	13.6	19.4	6.7
95 <sup>th</sup> %tile	6.3	12.5	18.5	26.1	10.2
99 <sup>th</sup> %tile	12.8	22.3	32.0	46.2	20.3
<i>N</i>	8,539,830	2,910,234	1,117,680	320,085	12,887,829

*Note:* This table reports daily usage statistics (of the *subscriber-day* usage distribution) for the four Internet service plans and entire sample.

\$58.89 per month for a 22 Mbps downstream connection with a 267 GB usage allowance. The maximum offered speeds and allowances are consistent with those offered in the US, but few subscribers choose them (as we have observed in the data of other ISPs, too). Exceeding the usage allowance is rare in this sample, only 1.6% of *subscriber-month* observations are over. This rate of overages is notably lower than the approximately 10% rate reported in Nevo et al. (2016), and is largely due to the recent substantial increase in allowances.

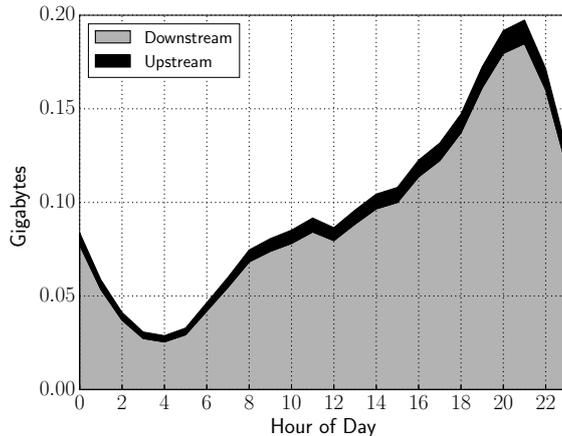
Subscribers on more expensive Internet plans use more data on average. In Table 1, we observe the daily usage distributions of higher tiers dominate those of lower tiers. This suggests differences in usage between tiers is at least partially driven by preferences for larger usage allowances and, possibly, faster speeds. While the differences in subscriber usage may be stark – for example, Tier 4 subscribers use 485% more data on average than Tier 1 subscribers – these extreme subscribers represent a only a modest percentage of the subscriber base. Tier 4 subscribers only account for 2.5% of the sample. Over 90% of *subscriber-day* observations are from Tiers 1 and 2.

## 2.2 Temporal Patterns in Usage

Hourly average usage in Figure 2 follows a cyclical pattern of maximum usage around 9PM and minimum usage around 4AM. This pattern is similar to what is found in Malone et al. (2014) and Nevo et al. (2016) with IPDR data from 2012. Usage during the 9PM peak hour is about 0.2 GB, over four times greater than the day’s trough. Throughout this analysis, we will refer to 6PM–11PM as *peak hours* and the rest of the day as *off-peak hours*.

About 40% of daily usage occurs during peak hours, as shown in panel (a) of Figure 5, with the 9PM hour accounting for just over 8% of a day’s usage. In panel (b), we present the proportion of usage by hour for different deciles of the total usage distribution. The

Figure 2: *Average Usage by Hour and Direction*



*Note:* This figure presents average hourly downstream and upstream usage in gigabytes.

general shape of panel (a) holds for subscribers, regardless of their overall usage level, with the heaviest subscribers (the 10th decile) having only a slightly flatter profile across the day than the others.

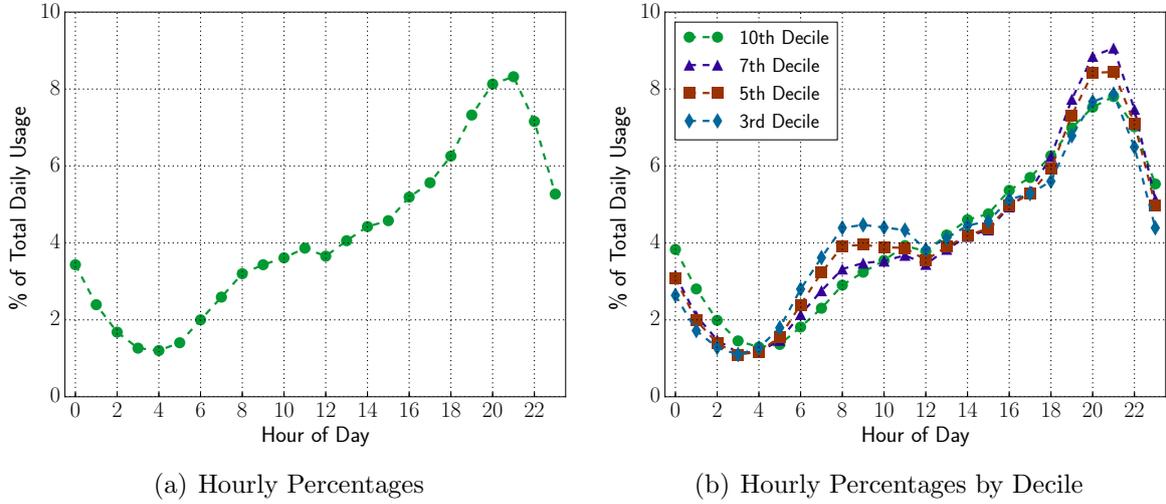
Together, Figures 2 and 5 show a strong and consistent pattern in usage across the day. This pattern suggests ISPs must invest enough in their network to meet demand or the network will become poor and unreliable. One unique feature of our data is we observe numerous periods of excess demand placed on the network that result in congestion. Additionally, we also observe the ISP make substantial investments to increase the capacity and improve the quality of their core network. The behavioral response of subscribers to variation in congestion is of primary interest to our analysis. We next discuss measures of congestion, and behavioral responses to congestion–mitigation efforts by the ISP.

### 2.3 Network Congestion and Packet Loss

Network congestion occurs when subscriber demand exceeds some capacity constraint on the network. During congested periods, subscribers may find that websites fail to load or online video buffers multiple times. There are two ways to measure congestion in our data. One is through hourly average node utilization. The node being the primary bottleneck in the “last mile” of an ISP’s network. The second being the hourly proportion of packets dropped/delayed, which we, and others, refer to as *packet loss*.

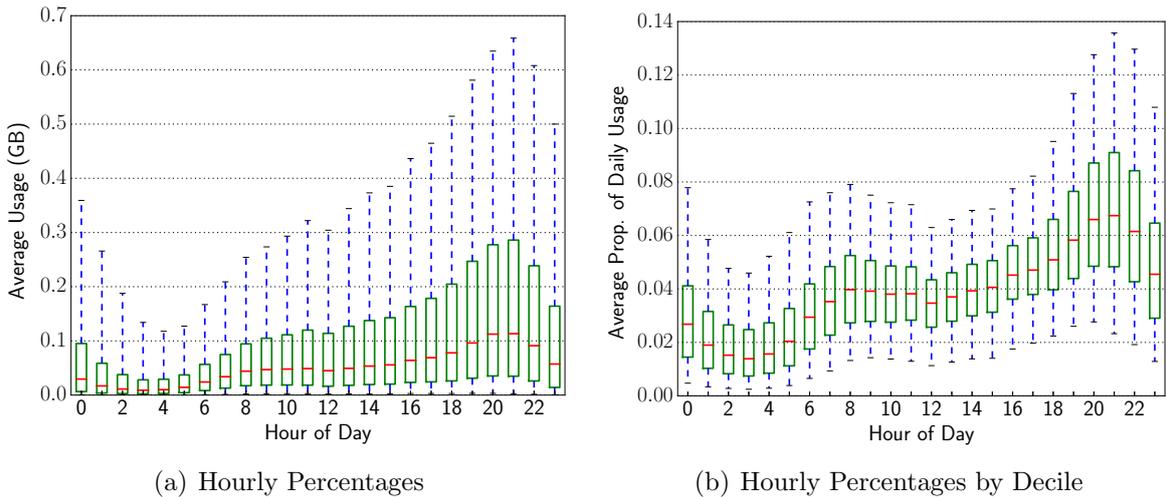
One advantage of hourly packet loss over node utilization is that packet loss is an individual measure instead of an aggregate one. Even when a node is highly utilized, some subscribers may have a normal experience over the hour. Packet loss occurs when data is undeliverable to a subscriber because current network delivery queues are full. Depending on if the data are dropped or delayed, the subscriber’s computer may have to request the data again, further increasing the time of delivery. Packet loss is more

Figure 3: *Statistics of Usage as a Percentage of Daily Total*



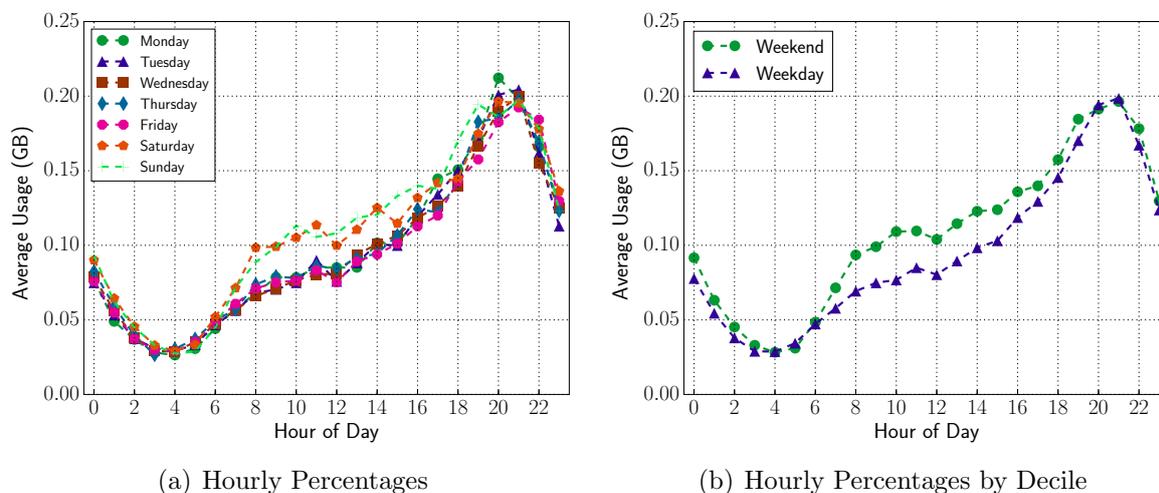
*Note:* This figure presents two figures related to how daily usage is proportionally distributed across the day. In panel (a), aggregate hourly percentages are reported for the entire sample. In panel (b), we conditionally report these hourly percentages for deciles 3, 5, 7, and 10. These deciles are calculated using total usage across the entire panel. Each series sums to 100%.

Figure 4: *Variation Across Households in Temporal Usage*



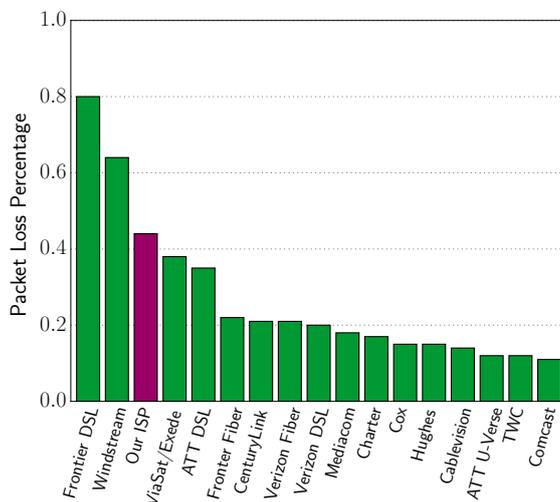
*Note:* This figure presents two figures related to how daily usage is proportionally distributed across the day. In panel (a), aggregate hourly percentages are reported for the entire sample. In panel (b), we conditionally report these hourly percentages for deciles 3, 5, 7, and 10. These deciles are calculated using total usage across the entire panel. Each series sums to 100%.

Figure 5: *Variation Across Households in Temporal Usage*



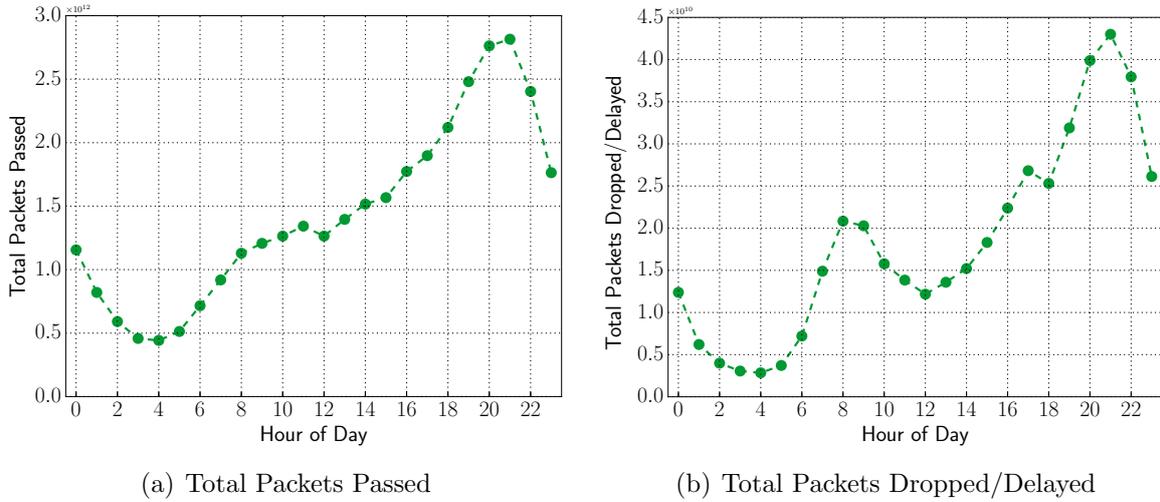
*Note:* This figure presents two figures related to how daily usage is proportionally distributed across the day. In panel (a), aggregate hourly percentages are reported for the entire sample. In panel (b), we conditionally report these hourly percentages for deciles 3, 5, 7, and 10. These deciles are calculated using total usage across the entire panel. Each series sums to 100%.

Figure 6: *Industry Statistics on Packet Loss from FCC's 2015 Report*



*Note:* This Figure is a reproduction from the FCC's 2015 Measuring Broadband America Fixed Report (<https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-2015>), which we have modified to include statistics from our ISP's sample. From the FCC report, it is not clear exactly how their statistics are calculated, but our personal experience with SamKnows data suggests the statistics are the average of hourly packet loss percentages from a specific test conducted by the modem. We calculate the average for "our ISP" according to this methodology. If we aggregate to a daily level, the average is 0.9%.

Figure 7: *Hourly Aggregate Packet Statistics*



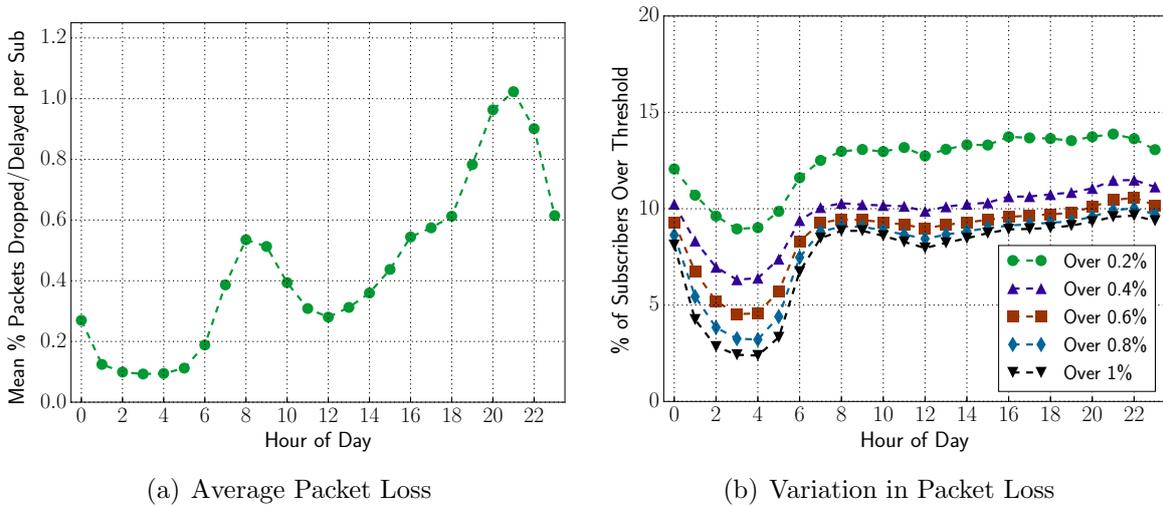
*Note:* These figures present hourly packet statistics. In panel (a), aggregate hourly totals of packets passed are reported. In panel (b), the total number of packets dropped/delayed are reported. These hourly totals are taken across the entire length of the panel.

likely to occur when nodes are highly utilized, which we observe in our data.<sup>13</sup> We only observe node utilization at the hourly level, too, which may not be granular enough to accurately reflect a subscriber’s experience within an hour. However, the performance of the network at the instant the subscriber sends and receives packets will be reflected in subscriber-specific hourly packet loss measures.

As part of the FCC’s efforts to monitor the quality of broadband networks, it produces an annual report on the state of broadband networks titled “Measuring Broadband America Fixed Broadband Report”. In the 2015 version, the FCC includes analysis of data from special (SamKnows) modems, which are distributed across numerous ISP networks. In addition to serving their normal function, these modems conduct hourly tests to measure the performance of the network. One of these tests seeks to measure packet loss by performing an FTP transfer to designated servers located throughout the US. Figure 8 of their report, which we have reproduced in Figure 6, presents statistics on packet loss from each ISP’s network. While it is not clear how the values in their Figure 8 are calculated, our experience working with identical data suggests the reported statistics are average hourly packet loss, where the average is taken across subscribers and hours of the day. In our reproduction of the FCC figure, we calculate the same statistic for our ISP. This particular measure of network performance would rate our ISP as the third worst across all types of networks in the FCC data (DSL, cable, fiber, and satellite). If you calculate the overall percentage of daily packets lost, rather than the average of hourly percentages, our ISP would be the worst at around 0.9% of all packets lost. By

<sup>13</sup>The two measures, hourly utilization and packet loss, have a correlation coefficient equal to 0.164.

Figure 8: Average Hourly Subscriber Packet Loss



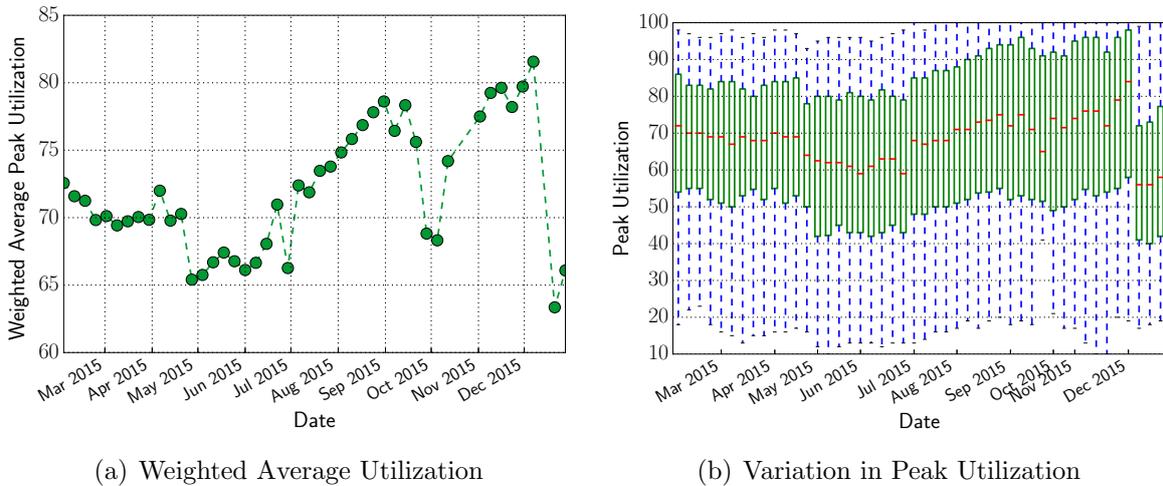
*Note:* These figures report statistics of average hourly packet loss. For each subscriber in the sample, we average hourly packet loss across the panel to generate these figures. In panel (a), we report average hourly packet loss for all subscribers. In panel (b), reports the percentage of average packet loss that is over various thresholds. For each subscriber in the sample, we average hourly packet loss across the panel. The percentage of hourly observations over 0.2%, 0.4%, 0.6%, 0.8%, and 1% are shown in the figure.

either definition, our sample represents a great opportunity to study the importance of congestion in broadband networks.

We now provide a more detailed analysis of the measures of congestion we have in our sample. More packets are passed during peak hours when usage is highest, as shown in panel (a) of Figure 7. This relationship follows from how requests are made online. Whenever a subscriber requests a website, a file to download, or a video to stream on the Internet, packets are sent between the subscriber’s computer and the item’s location. We also observe the highest frequency of dropped/delayed packets during peak hours in panel (b). The reasoning is twofold. First, there are more packets passed during these hours. Second, peak hours are when node utilization is highest.

At the subscriber level of observation, the distribution of hourly packet loss is highly skewed. For example, in panel (a) of Figure 8, average packet loss is around 1% at 9PM. However, from panel (b), we find over 85% experience less than 0.2% packet loss. Therefore, the majority of people experience little packet loss over the day, but in some cases, packet loss is very severe. The effects of packet loss on customer experience can be variable, too. For example, when watching a streaming video, 0.5% of packet loss may be acceptable for the video to finish. However, if someone is browsing a website, dropping a single packet could be the difference in a website failing to load correctly. This is important from a modeling standpoint, as we provide a flexible framework to estimate a rich distribution of tastes, which accounts for heterogeneity in the types of content the

Figure 9: *Weekly Node Utilization Statistics*



*Note:* Panel (a) plots the weighted average of peak utilization by week. The weights in this case are the number of people on the node. Panel (b) plots the weekly variation in peak utilization. The green box is the IQR, the red line is the median, and the blue dashed lines extend to the 5th and 95th percentiles.

individual prefers to consume.

Panel (b) of Figure 8 better captures the right-tail of the packet loss distribution. In this Figure, we present the the percentage of subscribers that are over various packet loss thresholds by hour. Notice in the early morning, when packet loss is lowest, about 3% of subscribers still experience about 1% packet loss on average, compared to the day’s maximum of 10% during peak hours. Interestingly, after 8AM the percentage of subscribers exceeding each threshold remain fairly constant over the remainder of the day.

## 2.4 Evolution of Network Quality

Since our data span a ten-month window, we observe changes in the overall quality of the network that, given the correlation between node utilization and packet loss, would improve packet loss and the network state. In panel (a) of Figure 9, the weighted average of peak node utilization is plotted for each week in our panel. Not only is there variation across the year, but there are distinct drops in May, September, and December where the ISP improved node capacity. These changes are also noticeable in how median peak utilization varies in panel (b). The dashed whiskers in panel (b) represent the 5th and 95th percentiles of peak usage, where even during these network events the variation within a week is unaffected.

One way an ISP can alleviate congestion on a node is to perform a *node split*. This is just one option available to an ISP – an ISP can use other hardware, software, and licensing methods to change the capacity of and bandwidth made available to a node. An example of a node split is for the operator to take a node and split its subscribers

Table 2: *Changes in Node Utilization and Packet Loss After Node Split*

	<i>Before</i>	<i>After</i>	<i>Diff</i>	<i>% Change</i>
Hourly Utilization	49%	34%	-15%	-31%
Max Hourly Utilization	87%	62%	-25%	-29%
Hourly Packet Loss	0.11%	0.08%	-0.03%	-27%
Max Hourly Packet Loss	1.0%	0.61%	-0.39%	-39%

*Note:* This table reports how the averages of node utilization and packet loss compare before and after the node split. 7 days of data is taken from before and after the node split date to calculate means. These averages are at the node level of observation and are weighted by the number of people on the node.

Table 3: *Changes in Daily Usage After Node Split*

	<i>Before</i>	<i>After</i>	<i>Diff</i>	<i>% Change</i>
Off-Peak Usage	1.51 GB	1.57 GB	0.06 GB	4.0%
Peak Usage	1.04 GB	1.16 GB	0.12 GB	12.0%
Total Daily Usage	2.55 GB	2.73 GB	0.18 GB	7.1%

*Note:* This table reports how subscriber behavior changed around a node split. 7 days of daily usage is taken from before and after the node split date to calculate means. This table summarizes usage for 2,627 subscribers over 5 node splits.

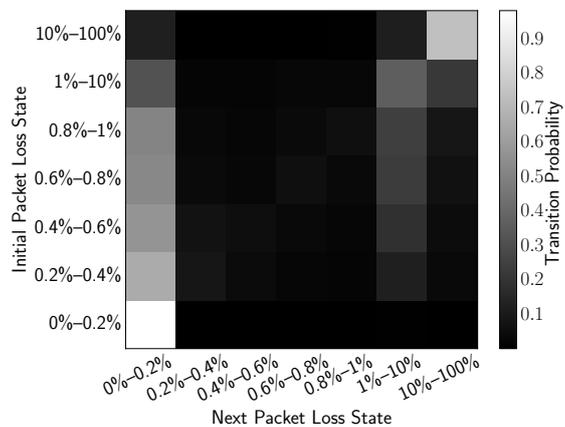
across two new nodes. When such a change is made, the network state for the affected subscribers should be improved since there are half as many subscribers using the same node. If subscriber behavior is responsive to such changes in network quality, we would expect an increase in usage. Note that the increase in usage could come from a change in the subscriber himself, or bandwidth adaptive applications becoming more responsive.

There are 5 distinct node splits in the data, whereby a group of subscribers is clearly split over two new nodes. Changes in network conditions are summarized in Table 2 and subscriber usage in Table 3. We do see improvements in the average network state with decreases in both utilization and packet loss. Maximum hourly node utilization falls by 29% and maximum hourly packet loss falls by 39%. Over this same period, we find a 7.1% increase in daily usage. Peak usage increases relatively more (12%) than off-peak usage (4.0%). This suggests that there is some degree of unmet demand prior to the node split that is now able to be realized.

In Table 4 and Figure 10, packet loss is split into seven bins that are used to study how persistent packet loss is day-to-day; the values in the heat map are the same as in the table. From these transition probabilities, there are a couple of notable takeaways. First, if a subscriber’s peak packet loss is poor one day, there is a high probability it will be better the next day. Second, if a subscriber does end up in the worst packet loss state, they are most likely to be in a poor state the next day. Third, the vast majority of subscribers experience low packet loss and will experience low packet loss tomorrow.

For the model, we use the transition matrix in Table 4 to estimate the frequencies

Figure 10: *Heatmap of Peak Packet Loss Transitions*



*Note:*

Table 4: *Transition Matrix of Peak Packet Loss*

Initial State	Next State						
	$0-0.2$	$0.2-0.4$	$0.4-0.6$	$0.6-0.8$	$0.8-1$	$1-10$	$10-100$
$0-0.2$	0.984	0.002	0.001	0.001	0.001	0.006	0.004
$0.2-0.4$	0.662	0.086	0.044	0.027	0.021	0.124	0.037
$0.4-0.6$	0.570	0.074	0.055	0.037	0.027	0.186	0.051
$0.6-0.8$	0.526	0.041	0.031	0.062	0.039	0.235	0.066
$0.8-1$	0.511	0.032	0.026	0.042	0.059	0.244	0.087
$1-10$	0.316	0.023	0.020	0.029	0.029	0.364	0.218
$10-100$	0.122	0.004	0.003	0.005	0.005	0.119	0.741

*Note:* This table reports probabilities of *peak hour-day* packet loss transitions at the subscriber level of observation. Each bin is of the form  $(x\%, y\%]$  and represent a range of packet loss. The first bin includes 0% packet loss, too.

of transition between packet loss, or network congestion, states. Below in the model discussion, this will be  $G_\psi$ . This matrix will be used to solve the model. For the estimation procedure, all we need are *day-hour* observations of daily consumption and the observed peak packet loss state for each account in the sample.

### 3 Model

Our model builds on the model of Nevo et al. (2016). Like Nevo et al. (2016), we assume a finite horizon, that a subscriber’s discount rate is  $\beta$ , and that a subscriber makes a consumption decision each period on his optimally chosen plan. The primary difference between the two models is we include network congestion and allow for it to impact subscribers’ plan and consumption choices.

Given that our focus is on the role of congestion, we limit our sample to only subscribers who never switched plans over the duration of the panel. This does not affect our analysis for two reasons. First, service plans were upgraded shortly before our sample, and, second, about 90% of subscribers made no changes during our period of observation. Allowing for plan switching introduces a dependency across billing cycles in the dynamic problem, and by not modeling it the computational burden of solving the model is reduced.<sup>14</sup>

#### 3.1 Subscriber Utility From Content

Subscribers derive utility from consumption of content. Each day of a billing cycle,  $t = 1 \dots T$ , a subscriber chooses the amount of content to consume on their chosen service plan,  $k = 1, \dots, K$ . Plans are characterized by a provisioned speed content is delivered,  $s_k$ , by a usage allowance,  $\bar{C}_k$ , by a fixed fee  $F_k$  that pays for all usage up to the allowance and by an overage price,  $p_k$ , per GB of usage in excess of the allowance. The menu of plans, and the characteristics of each, are fixed.<sup>15</sup> The provisioned speed is impacted by the state of the network,  $\psi$ , which changes daily due to variation in congestion and frequent network upgrades. We assume this evolution follows a first-order Markov process,  $G_\psi$ . Estimates of this process are presented in Table 4.

Utility from content is additively separable over all days in the billing cycle, and across billing cycles.<sup>16</sup> Let daily consumption of content be denoted by  $c$ . The utility for

---

<sup>14</sup>We explore the impact of relaxing this assumption in the Appendix by permitting customers to switch plans once at the beginning of each billing cycle, and including the excluded customers in our sample. We find little impact on our results.

<sup>15</sup>Plans were changed months prior to our sample, but unchanged during our sample, and the ISP had no plans to change them in the months after our sample ends.

<sup>16</sup>In this way, we assume content with a similar marginal utility is generated each day or constantly refreshed. This may not be the case for a subscriber who has not previously had access to the Internet. Below we will assume decreasing marginal utility within a time period, but additive across periods.

a subscriber of type  $h$  on plan  $k$  is given by

$$u_{hk}(c, \psi, v) = v \left( \frac{c^{1-\alpha_h}}{1-\alpha_h} \right) - c \left( \frac{\kappa_h}{\ln(\psi s_k)} \right).$$

The first term captures the subscriber’s utility from consuming the content. Marginal utility is declining, as we expect the first of any activity (email, web browsing, video, etc.) to bring higher marginal utility than subsequent usage. The convexity of the utility function is also quite flexible, nesting everything between log ( $\alpha_h \rightarrow 1$ ) and linear ( $\alpha_h = 0$ ). This leads to a straightforward link between  $\alpha_h$  and the price elasticity of demand, such that  $\alpha_h$  is the elasticity with respect to the entire cost associated with consuming content, both monetary and non-monetary. Uncertainty in utility from consumption of content is introduced by a time-varying shock,  $v$ , which is realized on the day the consumption decision is made. We assume that  $v$  is independently and identically distributed according to a log normal distribution with parameters,  $\mu_h^v$  and  $\sigma_h^v$ , for each type,  $h$ .

The second term captures the subscriber’s non-monetary cost of consuming content. This cost,  $\frac{\kappa_h}{\psi \ln(s_k)}$ , is time-varying and subscriber-specific. The  $\kappa_h > 0$  parameter captures both a subscriber’s *preference for speed* and the *waiting cost of transferring content*, which depends on the plan’s provisioned speed and the state of the network. Importantly, for any finite speed, this specification implies that each subscriber type has a satiation point even in the absence of overage charges. Thus, our specification of this cost departs slightly from that of Nevo et al. (2016), by not including an additive fixed value ( $\kappa_1$  in their model) that interacts with speed. This parameter is only weakly identified, as the limiting case with unbounded speed that would fully reveal this cost does not occur in our data. The interaction of the network state and speed captures the way in which this operator has chosen to ration bandwidth during congestion, as a proportional degradation of the provisioned speed.<sup>17</sup>

### 3.2 Optimal Usage

The observability of the network state and upgrade of plan features prior to our sample, which limits plan switching and permits focusing our analysis on the approximately 90% of consumers enrolled on a single plan the entire sample period, simplifies the characterization of optimal usage. Specifically, like Nevo et al. (2015), each consumer must solve a finite-horizon dynamic programming problem within each billing cycle. For a subscriber on plan  $k$ , we denote the amount of his unused usage allowance, on day  $t$  of the billing cycle, as  $\bar{C}_{kt} \equiv \text{Max}\{\bar{C}_k - C_{t-1}, 0\}$ , where  $C_{t-1}$  is cumulative usage up until day  $t$ . Similarly, denote day- $t$  overage as  $\mathcal{O}_{tk}(c_t) \equiv \text{Max}\{c_t - \bar{C}_{kt}, 0\}$ .

In the last day of the billing cycle ( $T$ ), the subscriber faces no intertemporal tradeoffs

---

<sup>17</sup>Our discussions with network engineers suggest this rationing rule can easily be altered to ration capacity differently during times of congestion.

and solves a static optimization problem, conditional on his cumulative usage  $C_{T-1}$  and the realization of the preference shock,  $v_T$ . Once  $v_T$  is realized, subscribers who will not incur overage charges (i.e.,  $\bar{C}_{kT}$  is high) consume such that  $\frac{\partial u_h(c_t, y_t, v_t; k)}{\partial c_t} = 0$ . If  $\frac{\partial u_h(c_t, y_t, v_t; k)}{\partial c_t}$  at  $c_t = \bar{C}_{kT}$  is positive and less than  $p_k$ , then consuming the remaining allowance is optimal. For those subscribers above the allowance (i.e.,  $\bar{C}_{kT} = 0$ ) and a high realization of  $v_T$ , it is optimal to consume such that  $\frac{\partial u_h(c_t, y_t, v_t; k)}{\partial c_t} = p_k$ . Denote this optimal level of consumption in each scenario by  $c_{hkT}^*(C_{T-1}, v_T)$ . Given this optimal policy for consumption, utility in the terminal period is

$$V_{hkT}(C_{T-1}, \psi_T, v_T) = v_T \left( \frac{(c_{hkT}^*)^{1-\alpha_h}}{1-\alpha_h} \right) - c_{hkT}^* \left( \frac{\kappa_h}{\ln(\psi_t s_k)} \right) - p_k \mathcal{O}_{tk}(c_{hkT}^*).$$

For other days in the billing period,  $t < T$ , consumption increases cumulative consumption and alters the state, so the optimal policy for a subscriber must incorporate this. The optimal policy for any  $t < T$  can be expressed recursively such that for type  $h$  on plan  $k$

$$c_{hkt}^*(C_{t-1}, \psi_t, v_t) = \underset{c_t}{\operatorname{argmax}} \left\{ v_t \left( \frac{c_t^{1-\alpha_h}}{1-\alpha_h} \right) - c_t \left( \frac{\kappa_h}{\ln(\psi_t s_k)} \right) - p_k \mathcal{O}_{tk}(c_t) + E_\psi [V_{hk(t+1)}(C_{t-1} + c_t, \psi_{t+1})] \right\}.$$

Alternatively, defining the *shadow price* of consumption as

$$\tilde{p}_k(c_t, C_{t-1}, \psi_t) = \begin{cases} p_k & \text{if } \mathcal{O}_{tk}(c_t) > 0 \\ \frac{dE_\psi[V_{hk(t+1)}(C_{t-1} + c_t, \psi_{t+1})]}{dc_t} & \text{if } \mathcal{O}_{tk}(c_t) = 0. \end{cases}$$

the optimal consumption choice in period  $t$  satisfies

$$c_{hkt}^* = \left( \frac{v_t}{\frac{\kappa_h}{\ln(\psi_t s_k)} + \tilde{p}_k(c_{hkt}^*, C_{t-1}, \psi_t)} \right)^{\frac{1}{\alpha_h}}. \quad (1)$$

The relationship between  $\alpha_h$  and the price elasticity of usage is clear from Equation 1. A type with parameter  $\alpha_h$  has a usage elasticity equal to  $-\frac{1}{\alpha_h}$  with respect to the entire marginal cost of content,  $\frac{\kappa_h}{\ln(\psi_t s_k)} + \tilde{p}_k(c_t, C_{t-1}, \psi_t)$ .

The value function associated with the optimal usage policy is

$$V_{hkt}(C_{t-1}, \psi_t, v_t) = v_t \left( \frac{(c_{hkt}^*)^{1-\alpha_h}}{1-\alpha_h} \right) - c_{hkt}^* \left( \frac{\kappa_h}{\ln(\psi_t s_k)} \right) - p_k \mathcal{O}_t(c_{hkt}^*) + E_\psi [V_{hk(t+1)}(C_{t-1} + c_{hkt}^*, \psi_{t+1})]$$

for each 3-tuple,  $(C_{t-1}, \psi_t, v_t)$ . Then for all  $t < T$ , the expected continuation value is

$$E_\psi [V_{hkt}(C_{t-1}, \psi_{t+1})] = \int_\psi \left( \int_v V_{hkt}(C_{t-1}, \psi_{t+1}, v_t) dG_v^h(v_t) \right) dG_\psi^h(\psi_{t+1}|\psi_t),$$

and the mean of a subscriber's usage at each observable state is

$$c_{hkt}^*(C_{t-1}, \psi_t) = \int_v c_{hkt}^*(C_{t-1}, \psi_t, v_t) dG_v^h(v_t). \quad (2)$$

The Markov process associated with the the solution to the dynamic program also implies a distribution for the time spent in each state,  $(t, C_{t-1}, \psi_t)$ , over a billing cycle,  $P_{hk_h^*m}(C_{m-1}, \psi_m)$ . This process along with expected consumption at each state form the basis of our estimation algorithm.

### 3.3 Plan Choice

We assume subscribers select plans to maximize expected utility, before observing any utility shocks, and remain on that plan during our sample. More precisely, we assume that the subscriber selects one of the offered plans,  $k \in \{1, \dots, K\}$ , or no plan,  $k = 0$ , such that

$$k_h^* = \underset{k \in \{0, 1, \dots, K\}}{\operatorname{argmax}} \{E[V_{hk1}(C_1 = 0, \psi_1)] - F_k\}.$$

The optimal plan,  $k_h^*$ , maximizes expected utility for the subscriber given the current state of the network and optimal usage decisions,  $E[V_{hk1}(C_1 = 0, \psi_1)]$ , net of the plan's fixed access fee,  $F_k$ . The outside option is normalized to have a utility of zero. Note, that we assume that there is no error, so consumers choose the plan that is optimal. Similar to Nevo et al. (2015), (admittedly weak) tests of optimal plan choice reveal that it is rare to observe a subscriber whose usage decisions are such that switching to an alternative plan would yield a lower total costs at no slower speeds. The weakness of this optimality test is due to the positive correlation between speed and usage allowances of the offered plans (see Figure 1). Our assumptions on plan choice are easily relaxed in theory, but introduce a substantial additional computational burden. Given the infrequency of both clear ex-post mistakes in choosing a plan and switching of plans, we believe this is a reasonable assumption for our sample.

## 4 Estimation

Our estimation approach is a panel-data modification of Fox et al. (2015), proposed in Nevo et al. (2015), which we refer to as fixed-grid fixed-effect (FGFE) least squares. The approach exploits the richness of panel data to build upon the fixed-grid random-effects (FGRE) approach of Fox et al. (2015) used by Nevo et al. (2015). In contrast

to the FGRE approaches, our FGFE approach permits identification of each subscriber's type, rather than just the distribution of types, and also allows consideration of moments from the model that are not non-linear in the type-specific population weights. This is advantageous for identification of the model and consideration of richer counterfactual exercises where knowledge of an individual's type is useful rather than just the distribution of types.

#### 4.1 Econometric Objective Function

For each individual,  $i = 1 \dots I$ , we have a time series of data,  $m = 1 \dots M$ , which captures usage at a daily frequency on an optimally chosen plan.<sup>18</sup> Thus, we have a daily time series for usage,  $(c_{i1}, c_{i2}, \dots, c_{iM})$ , for individual  $i$ , as well as the accompanying observable portion of the state,  $(t_m, C_{m-1}, \psi_m)$  for each  $m = 1 \dots M$ . From the solution to the model, for each type  $h$ , we store two moments associated with usage on the optimally chosen plan,  $c_{hk^*t}^*(C_{t-1}, \psi_t)$  and  $c_{hk^*t}^{*2}(C_{t-1}, \psi_t)$ , expected usage and the expectation of the square of usage at every observable state. Additionally, we calculate the probability of observing a type in a particular state,  $P_{hk^*t}(C_{t-1}, \psi_t)$ .

The goal of the estimation algorithm is to identify which of the  $H$  types' behavior from the model best match the behavior of each individual,  $i$ , over the panel of data. We use a least-squares criteria to compare fit, such that the type  $h$  that best matches to consumer  $i$  is given by

$$\hat{h}_i = \min_{\{h=1 \dots H\}} \left[ \sum_{m=1}^M \tilde{\mathbf{z}}'_{ih} \tilde{\mathbf{z}}_{ih} \right],$$

where

$$\tilde{\mathbf{z}}_{ih} = \begin{pmatrix} c_{im} - c_{hk^*m}^*(C_{m-1}, \psi_m) \\ c_{im}^2 - c_{hk^*m}^{*2}(C_{m-1}, \psi_m) \\ 1 - P_{hk^*m}(C_{m-1}, \psi_m) \end{pmatrix}.$$

This process is repeated for each  $i$ . Aggregating across the chosen types for each consumer,  $i$ , the population weights for each type,  $h$ , is then

$$\hat{\theta}_h = \frac{1}{I} \sum_{i=1}^I \mathbf{1}[\hat{h}_i = h].$$

There are a numerous advantages of the FGFE approach in panel-data applications

---

<sup>18</sup>We drop the small fraction of subscribers, less than 2%, for which we do not observe a complete time series.

like ours. First, even compared to the constrained convex optimization problem in the FGRE approach of Fox et al (2015), there can be computational advantages introduced by only searching over a fixed grid of types. Second, importantly, the FGFE approach does not require that the moments used in estimation be linear in the type-specific weights. In the FGRE approaches, this linearity is necessary or the problem becomes a constrained nonlinear optimization problem that is intractable with even a moderate number of types. These richer moments can be particularly helpful in identification, as Nevo and Williams (2016) show.

Another advantage of the FGFE approach is that it fully characterizes the discrete distribution of types, but in contrast to FGFE demand models like Fox et al (2015) and Nevo et al (2015), the mapping between an individual and a type is preserved ( $\hat{h}_i$ ). The panel data eliminates the need to aggregate across consumers to form moments, and this permits an individual's type to be inferred, rather than just the distribution of types in the population. In many applications in Industrial Organization, inferring an individual's type rather than only the distribution of types is useful. From the firms' perspective, this may permit different forms of discrimination (third-degree rather than second degree) to be implemented through targeted offerings. Knowledge of each individual's type can also permit a decomposition of the parameters via the minimum-distance procedure of Chamberlain (1982) when observable characteristics of the individual are available, as is done in Nevo (2001). For example, one can regress the parameters describing an individual's type,  $(\mu_{\hat{h}_i}^v, \sigma_{\hat{h}_i}^v, \alpha_{\hat{h}_i}, \kappa_{\hat{h}_i})$ , on a vector of the individual's observed characteristic to decompose the parameter into observable and unobservable determinants of preferences. This may be particularly useful in labor and health applications where a rich set of observable characteristics are often available.

## 4.2 Identification

Identification of our model closely follows the discussion in Nevo et al. (2015). There are a few important differences, each simplifying and improving identification. First, we eliminate the  $\kappa_{1h}$  parameter that led to a satiation point for usage even when speed was unbounded. While we observe higher speeds than Nevo et al. (2015), this dimension to the type space is not needed because the limiting case is clearly not in our data, and given the additional computational complexity of our model, eliminating it permits us to consider a denser grid over the other parameters. Second, like Nevo et al. (2015), usage and plan choices are strong sources of identification. Plan choice can be thought of as assigning each type to a plan and putting a uniform prior over the types on each plan, while the usage moments can then distinguish between the types choosing a plan. The flexibility of the FGFE approach is also important here, as we are able to consider richer usage moments, because we are not restricted to moments that preserve linearity in the type weights. This is how we're able to consider the first and second conditional

moments of usage, in contrast to Nevo et al. (2015), which only uses the unconditional first moment. Finally, we also have an additional source of variation in the price of usage that can help identify a type. Like Nevo et al. (2015), usage-based pricing is particularly helpful, as we observe a large number of marginal decisions by each consumer, weighing the benefit of consuming more content against the increase in the probability of overages (i.e., the shadow price of usage). This variation is helpful for pinning down the primary determinant of an individual’s elasticity of demand,  $\alpha_{\widehat{h}_i}$ . However, in addition to this price variation introduced by the nonlinear pricing, we also have extensive variation in the network state, which shifts the cost of consuming content. This is helpful in pinning down an individual’s preference for speed,  $\kappa_{\widehat{h}_i}$ , which would otherwise largely be identified by plan choice alone.

## 5 Results

We present our estimation results in two parts. First, we report our estimates of the types distributions. Next, we discuss the results of a counterfactual exercise that measures the value to consumers of eliminating all congestion on the network.

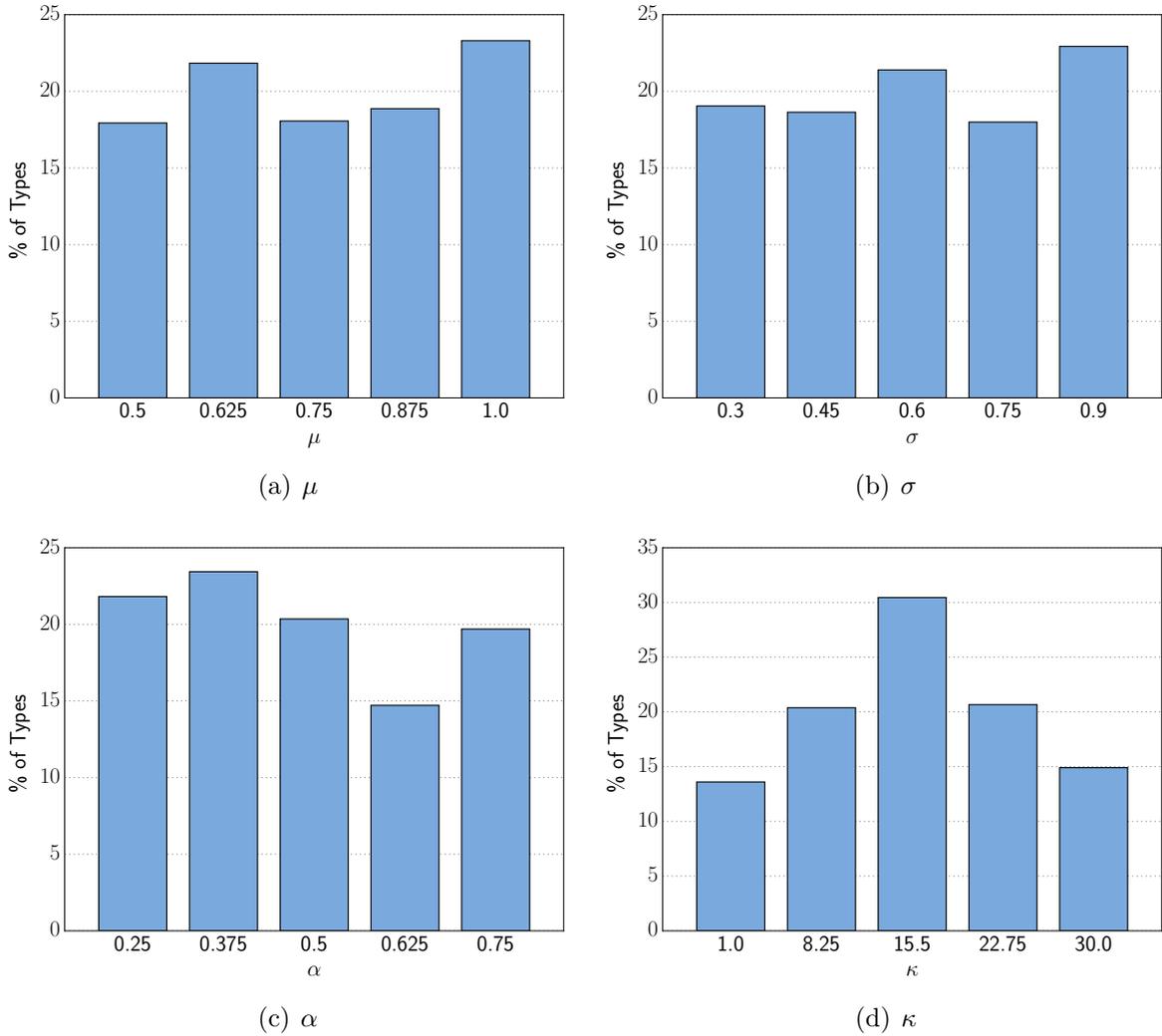
### 5.1 Type Distribution Estimates

We estimate a weight greater than 0.01% for 164 types. That is, 164 different types  $h$  were chosen for at least one subscriber ( $i$ ), or 164 different  $\widehat{h}_i$  were chosen among all possible types. Conditional on being chosen, we find the weights are distributed rather uniformly among the types. This is in contrast to Nevo et al. (2016), which finds a concentrated distribution of types. Their most common type accounted for 28% of the total mass, the top 5 types accounted for 65%, the top 10 for 78% and the top 20 for 90%. Nevo and Williams (2016) show this difference in the concentration of the types tend to be largely due to the difference between the FGFE employed here and the FGRE effects approach employed by Nevo et al. (2016).

Interestingly, the much larger number of types we estimate here is almost exclusively due to the most expensive plan, which accounts for 110 of the 164 positive types. On the most expensive plan, there is a wide variety of behavior that must be explained. We observe many low usage subscribers, which when optimal plan choice is assumed, can only be rationalized by a type with an intense preference for speed – think of an individual that video conferences or occasionally downloads very large files, and wants the applications used to perform seamlessly. Similarly, we have many individuals that desire a large allowance, but have a less-intense preference for speed.

Figure 11 presents the marginal distributions for each of the parameters,  $(\mu_h, \sigma_h, \kappa_h, \alpha_h)$ . Interestingly, the type distribution for each parameter is quite uniform, or non-normal, across the support. This contrasts the lumpy distributions recovered by Nevo et al. (2016). Therefore, the mean, median, and mode of the four parameters, which are re-

Figure 11: *Estimated Marginal Distributions of Model Parameters*



*Note:* The figures are the estimated marginal distributions of the four parameters in our model. Within each panel, the sum of all five bars in each figure total 100%.

Table 5: *Descriptive Statistics for Types*

	<i>Mean</i>	<i>Median</i>	<i>Mode</i>
$\mu_h$	0.759	0.75	1
$\sigma_h$	0.610	0.6	0.9
$\kappa_h$	15.709	15.5	15.5
$\alpha_h$	0.483	0.5	0.375

*Note:* This table reports descriptive statistics of the type distribution: mean, median, and mode.

ported in Table 5, are quite similar. For each of the parameters, the mean, median, and mode are within 10% of one another.

The estimated joint distributions are much more irregular, neither uniform or normal. These joint distributions for each combination of the parameters, six in total, are presented in Figure 12. Like Nevo et al. (2016), the joint distributions are multi-peaked and vary considerably by the pair of parameters considered. This demonstrates the importance of the flexibility of our estimation approach, which allows for free correlations between each pair of parameters rather than the zero covariance often assumed in structural econometric applications. This flexibility is reflected in the fit of the model. For all plans, the correlation between the empirical moments and the fitted moments is above 90%. The model also fits patterns in the data not explicitly used in estimation, similar to those reported in Nevo et al. (2016).

## 5.2 Value of Eliminating Congestion

In this counterfactual exercise, we measure the value to subscribers from eliminating network congestion, or the case where provisioned speeds are always realized to subscribers. This exercise is import because it illustrates the value, or lack thereof, of many types of core network improvements that go towards reducing congestion and better meeting demand.

By eliminating congestion, we estimate an increase in consumer surplus of 14%, as shown in Table 6. Daily usage increases from 2.2 GB/day to 2.5 GB/day, or about an additional 9 GB over a 30-day billing cycle. We also observe many subscribers downgrading plans as a result of the improved network state. Since the speeds of a cheaper plan are now guaranteed, subscribers with a stronger preference for speed over usage may be better served on the cheaper plan. This movement to cheaper plans lowers average revenue for the ISP, but the increase in consumer surplus is large enough to offset this drop: \$4.24 drop in revenue and \$8.90 increase in consumer surplus for a net difference of \$4.66. Since subscribers are receiving speeds roughly 19% faster, we estimate a subscriber values each additional Mbps of realized speed at \$2.87.

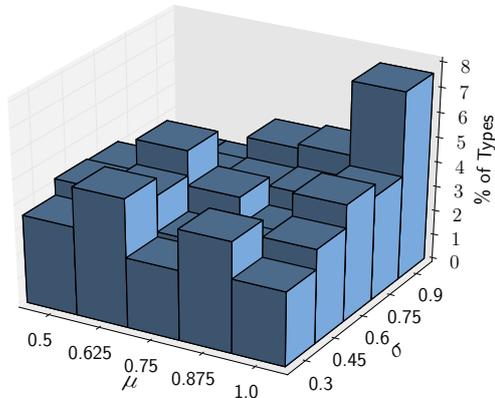
## 6 Conclusion

We estimate demand for residential broadband using a 10-month panel of hourly subscriber usage and network conditions. The key feature of our model is the incorporation of network congestion and allowing it to affect a subscriber's daily consumption decision.<sup>19</sup> There are three sources of variation we exploit in our data. First, we use (shadow) price variation that results from the structure of usage-based pricing's three-part tariff. Second, we use cross-sectional variation in packet loss, our measure of network congestion,

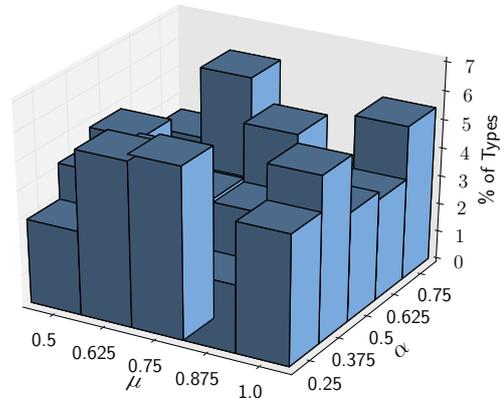
---

<sup>19</sup>This differs from previous research such as Nevo et al. (2016).

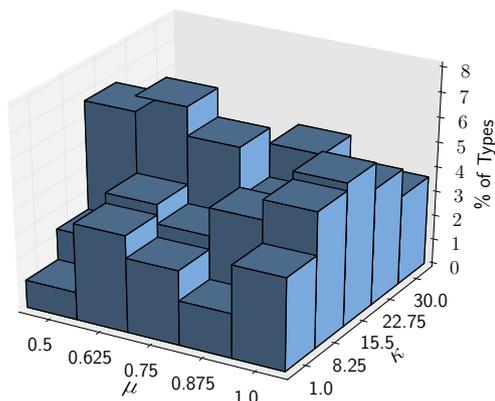
Figure 12: *Estimated Joint Distributions of Model Parameters*



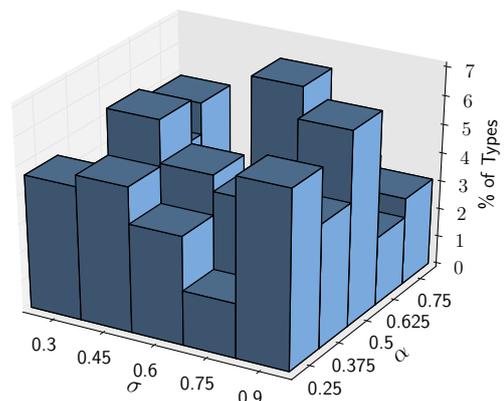
(a)  $\mu$  and  $\sigma$



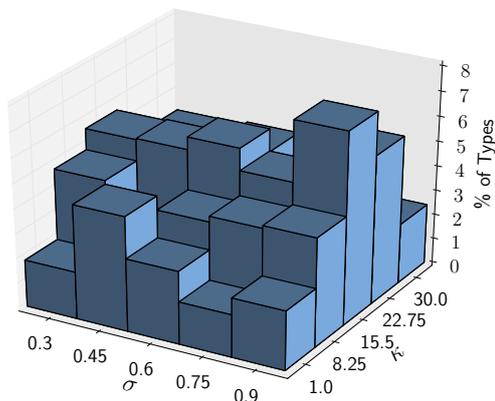
(b)  $\mu$  and  $\alpha$



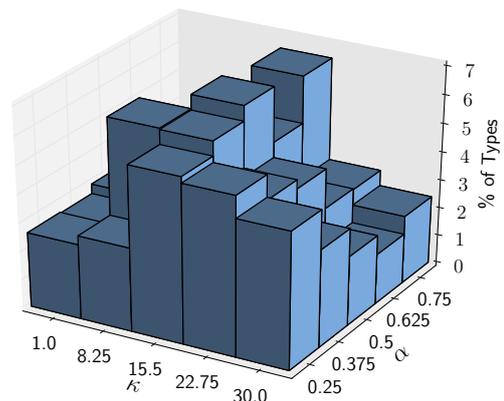
(c)  $\mu$  and  $\kappa$



(d)  $\sigma$  and  $\alpha$



(e)  $\sigma$  and  $\kappa$



(f)  $\kappa$  and  $\alpha$

*Note:* The figures are the estimated joint distributions of the four parameters in our model. Within each panel, the sum of all twenty-five bars in each figure total 100%.

Table 6: *Counterfactual Results from Eliminating Network Congestion*

Usage and Surplus	Current Offerings	
	<i>With Congestion</i>	<i>No Congestion</i>
Daily Usage (GB)	2.2 GB	2.5 GB
Provisioned Speed (Mbps)	22.3	19.8
Realized Avg. Speed (Mbps)	16.7	19.8
Consumer Surplus (\$)	65.80	74.70
Revenue (\$)	58.58	54.34

*Note:* This table reports results from the counterfactual exercise of eliminating network congestion, i.e., provisioned speeds are always delivered to subscribers.

across subscribers. Third, our ISP invested in the core network several times throughout 2015, creating times series variation in the overall quality of the network.

Our demand estimates are used to measure the value to subscribers from eliminating network congestion. We find the improved network conditions encourage some subscribers to downgrade, but any loss in revenue is entirely offset by an increase in consumer surplus. Subscribers’ realized speeds increased by roughly 18% with each additional Mbps of speed being valued at roughly \$2.87.

There are several extensions to the basic model in this paper that can be explored in future versions. First, we could allow consumers to switch plans. In our sample, around 12% of subscribers make a plan change by either moving to a more expensive or cheaper plan; these subscribers are omitted from our original estimation. Under a usage-based pricing regime, we expect some consumers may upgrade to a plan with a larger usage allowance to account for their growing demand, while others may downgrade to better align their usage with a lower allowance or due to cost concerns.

Next, we could allow consumers to make two consumption choices each day: one for off-peak hours (12AM–5PM) and peak hours (6PM–11PM). This permits two additional pieces of analysis. First, we are able to study how congestion differentially affects usage during different times of the day. Since demand is greater during peak hours, we may expect subscribers to behave differently. We are also able to explore the welfare implications of peak-use pricing as an implementation of usage-based pricing. Large demand during peak hours drives higher service costs for the ISP during these hours. However, during off-peak hours, service is mostly costless since the network is less congested. For example, if only peak hour usage counted towards the usage allowance,<sup>20</sup> we may observe subscribers flattening their usage profile across the day by shifting peak hour usage to off-peak times.

Then, in related research, we believe there are many related topics. First, our hourly data usage is aggregated across all traffic types. Getting high-frequency data that are

<sup>20</sup>This type of usage-based pricing was common with the telephone companies when they offered unlimited minutes on nights and weekends.

disaggregated by application or traffic type would permit a more detailed analysis and understanding of how congestion differentially affects each type. Moreover, moving to a more granular level, say 5 to 15 minutes, would allow for a more exact understanding of the correlation between usage and congestion. Second, residential broadband and traditional linear television (TV) services are closely related and are often bundled together by ISPs. In future research, we hope to obtain linear TV data in conjunction with disaggregated high-frequency Internet usage by type to more completely explore how subscribers use broadband Internet. Specific relationships like the substitutability of linear TV and over-the-top video (OTTV) and how linear TV usage and network congestion are correlated could be explored with such a data set.

## References

- Dutz, Mark, Jonathan Orszag and Robert Willig (2009). “The Substantial Consumer Benefits of Broadband Connectivity for US Households.” *Internet Intervention Alliance Working Paper*.
- Edell, Richard and Pravin Varaiya (2002). *Providing Internet Access: What We Learn from INDEX*, volume Broadband: Should We Regulate High-Speed Internet Access? Brookings Institution.
- Goolsbee, Austan and Peter Klenow (2006). “Valuing Products by the Time Spent Using Them: An Application to the Internet.” *American Economic Review P&P*, 96(2): 108–113.
- Greenstein, Shane and Ryan McDevitt (2011). “The Broadband Bonus: Estimating Broadband Internet’s Economic Value.” *Telecommunications Policy*, 35(7): 617–632.
- Hitte, Loran and Prasanna Tambe (2007). “Broadband Adoption and Content Consumption.” *Information Economics and Policy*, 74(6): 1637–1673.
- Lambrecht, Anja, Katja Seim and Bernd Skiera (2007). “Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs.” *Marketing Science*, 26(5): 698–710.
- Malone, Jacob, Aviv Nevo and Jonathan Williams (2016). “A Snapshot of the Current State of Residential Broadband Networks.” *NET Institute Working Paper No. 15-06*.
- Malone, Jacob, John Turner and Jonathan Williams (2014). “Do Three-Part Tariffs Improve Efficiency in Residential Broadband Networks?” *Telecommunications Policy*, 38(11): 1035–1045.
- Nevo, Aviv, John Turner and Jonathan Williams (2016). “Usage-Based Pricing and Demand for Residential Broadband.” *Econometrica*, 84(2): 411–443.

Rosston, Gregory, Scott Savage and Bradley Wimmer (2013). “Effect of Network Unbundling on Retail Price: Evidence from the Telecommunications Act of 1996.” *Journal of Law and Economics*, 56(2): 487–519.

Varian, Hal (2002). *The Demand for Bandwidth: Evidence from the INDEX Experiment*, volume Broadband: Should We Regulate High-Speed Internet Access? Brookings Institution.