Usage-Based Pricing and Demand for Residential Broadband *

Aviv Nevo†
Northwestern University

John L. Turner‡
University of Georgia

Jonathan W. Williams§
University of Georgia

September 2013

Abstract

The increasing use of the Internet creates a need to manage traffic while preserving equal treatment of content. We estimate demand for residential broadband, using high-frequency data from subscribers facing a three-part tariff, and use the estimates to study the welfare implications of usage-based pricing, a commonly offered solution to network congestion. The three-part tariff makes data usage during the billing cycle a dynamic problem; thus, generating variation in the (shadow) price of usage during the month. We provide evidence that subscribers respond to this variation, and use their dynamic decisions to estimate a flexible distribution of willingness to pay for different plan characteristics. Using these estimates, we show that usage-based pricing eliminates low-value traffic and improves overall welfare. Usage-based pricing might decrease consumer surplus, depending on what alternative is considered. Furthermore, we show that the costs associated with investment in fiber-optic networks (an alternative proposed to deal with congestion) are likely recoverable in some markets.

Keywords: Demand, Broadband, Dynamics, Usage-based Pricing.

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, Gautam Gowrisankaran, Shane Greenstein, Tom Holmes, Ron Reuss, and Scott Savage for insightful comments. Jim Metcalf provided expert IT support for this project. We gratefully acknowledge support from the NSF (Grants SES-1324851 and SES-1324717)

†Department of Economics, Northwestern University, nevo@northwestern.edu.

‡Department of Economics, University of Georgia, jlturner@uga.edu.

§Department of Economics, University of Georgia, jonwms@uga.edu.

1 Introduction

There are now about 210 million broadband Internet users in the US, up from roughly eight million in 2000. Taking advantage of an increasing variety of websites, devices and applications, these users spend an average of about thirty hours per month using the Internet (Rainie and Packel 2001; Nielsen 2012). The proliferation of online activities has led at times to congestion, and a need to allocate scarce bandwidth, especially in the "last mile" connecting subscribers to the Internet backbone. The providers of service in the last part of the connection, usually either telecommunications or cable companies, need to balance efficient pricing, which allocates the scarce resource, with preserving a principle that all data on the Internet should be treated equally, so-called "net neutrality."

One way to balance these forces is to rely on usage-based pricing, in its popular form a three-part tariff. Subscribers pay a monthly fee, which provides them a monthly data allowance. If they exceed the allowance they pay a price per Gigabyte (GB). This type of pricing is popular for broadband Internet service outside the US, and for cellular plans in the US. However, usage-based pricing has generated a policy discussion in the US when proposed as the standard for pricing broadband Internet service (OIAC 2013).[1] Much of the debate on the welfare implications of usage-based pricing has been theoretical (Mackie-Mason and Varian 1995; Bauer and Wildman 2012; Odlyzko et al. 2012), and has not been informed by data and empirical estimates.

In this paper, we address this policy debate. We estimate demand for broadband service and use the estimates to evaluate the welfare implications of usage-based pricing as well as other alternatives proposed to address network congestion. In particular, we compare the welfare from the usage-based pricing we observe in the data to a counterfactual where overage charges are eliminated, but plans are otherwise unchanged. Next, we compute welfare if Internet Service Providers (ISP) set revenue-maximizing prices for these unlimited plans. Finally, we evaluate welfare when only a single, high-speed plan is offered.

At the core of the paper is a unique data set that we have secured from a group of North American ISPs. In this paper we focus on data from one of these providers. The data include information on Internet usage by roughly 55,000 subscribers across multiple billing cycles during 2012, from different markets, facing different price schedules. For each household we know hour-by-hour data usage over the entire time they are observed. We also know plan-specific variables (speed, prices, etc.) for the plan the household is subscribed to and for the alternatives not

[1] Senator Ron Wyden also recently introduced the Data Cap Integrity Act Bill that seeks to place restrictions on when ISPs can use usage-based pricing (see http://www.wyden.senate.gov/news/press-releases/wyden-data-cap-legislation-will-protect-consumers-and-promote-innovation).

chosen. An important feature of the data is that the ISP we study currently has in place three-part tariff plans in addition to subscribers who are grandfathered in to unlimited plans (i.e., an overage price of zero).

A key challenge in estimating demand is finding price variation that can be used to recover subscribers' preferences. When facing a three-part tariff plan, the marginal price paid for usage is zero until the subscriber uses up their allowance. However, a forward-looking user realizes that the shadow price of usage is not zero. Indeed, the shadow price depends on how many days are left in the billing cycle and the fraction of the allowance already used. To exploit this variation, we build a dynamic model of utility-maximizing subscribers' inter-temporal decision making throughout a billing cycle.

The dynamic model allows us to take advantage of the high-frequency nature of our data and the variety of three-part tariffs offered to subscribers to estimate price elasticities. Specifically, we observe subscribers, at each point in the billing cycle, making decisions that weigh the value of consuming content today against an increased probability of incurring overage charges later. This type of variation traces out marginal utility. Also, selection into plans or choices of particular three-part tariffs reveal an average willingness to pay for content and a preference over the speed of one's connection. The ISP offers plans ranging from almost linear tariffs (i.e., very low usage allowances) to plans with allowances well over 100 GBs.

To start we provide some descriptive statistics from the rich data at our disposal. First, we present statistics on broadband usage and how it changes over time. From May 2011 to May 2012, the average (median) subscriber increased per-month usage from 23.1 (9.0) to 40.3 (20.3) GBs, an increase of 74.6% (125.5%). This rapid growth and the skewed nature of per-subscriber usage are often used as arguments in support of the need to control congestion with usage-based pricing. Second, we provide support for the behavioral assumptions we make in our model. We measure how many subscribers could have saved by choosing a different plan, holding the connection speed and usage fixed. In contrast to work studying other industries (Grubb and Osborne 2012; Grubb 2012; Handel 2013), we cannot reject that most subscribers choose what is ex-post the optimal plan. Third, we show that subscribers are responsive to the possibility of incurring overages: as the shadow price increases, subscribers decrease their usage. We believe that subscribers' response to the possibility of incurring overage charges and infrequency of ex-post mistakes are aided by the nearly real time information on usage provided by the ISP. Both findings support the use of a dynamic model that relies on optimal plan choice to recover preferences.

To estimate the model, we adapt the techniques of Ackerberg (2009), Bajari et al. (2007)

and Fox et al. (2011). The basic idea is intuitive. We solve the dynamic programming problem for a large number of subscriber types. The solution of the model is done once for each type, and can be done in parallel using a large number of processors. We then estimate the distribution of types by computing a weight for each of these types. We choose weights to match the empirical moments to those predicted by a weighted average of the optimal behavior of types. The weights give us the distribution of the various types, and can be estimated without relying on parametric distributional assumptions. We find that the estimated distribution looks nothing like continuous distributions usually assumed in empirical work, and that the fit deteriorates when the number of types is restricted, which is an alternative assumption often made.

We use the estimated distribution of preferences to calculate the willingness-to-pay for small changes to specific plan features, keeping plan choices and other features constant. We find the marginal willingness-to-pay for a one GB increase in usage allowance averages \$0.45 per month, which suggests the content that usage-based pricing is removing from the network is of relatively low value. Over 80% of subscribers would pay a positive amount to increase their usage allowances. We find the willingness-to-pay for a one Megabit per second (Mb/s) increase in connection speed is between \$0 to \$5.86 per month, with an average of \$1.76.

Next, we use the estimated type distribution to estimate expected demand for content assuming that the ISP drops all fixed fees and sells content for a flat price per GB. For a price of zero and with average download speed, we find per-subscriber usage would average about 71 GB per month and consumer surplus would average nearly \$280 per month. This highlights the tremendous potential value of Internet usage. We also find that average subscriber demand is inelastic for prices up to \$4 per GB. Intuitively, non-price costs of using the Internet, such as waiting for content to download, form a substantial part of the opportunity cost of spending time online and make subscribers relatively insensitive to small prices. When speed is significantly higher, waiting costs are lower and demand becomes more sensitive to the price. In addition, increases in speed lead to increases in usage that are comparably bigger than increases in willingness to pay.

We conduct a number of counterfactuals to explore the efficiency of usage-based pricing. The first counterfactual considers the welfare implications of usage-based pricing by eliminating overage charges while keeping unchanged the number of plans and their features, such as speed and fixed fees. We find that subscriber welfare is lower under usage-based pricing. Usage, and the cost it implies, is also lower and the ISP revenues are higher. Thus, the ISP is better off under usage-based pricing. Total welfare is higher under usage-based pricing, mostly due to the lower costs associated with 11% less traffic.[2] Interestingly, some subscribers actually *increase* usage under usage-based pricing. These subscribers choose a higher-speed plan under usage-based pricing, and since in our model higher speed reduces the waiting cost of consuming Internet content, this reduces their implied marginal price. So while the overall cost, or the average cost, of consuming content goes up under usage-based pricing for these subscribers, the marginal cost of consuming content actually goes down.

This calculation from the above counterfactual overestimates the harm to subscribers from usage-based pricing since we keep all plan features, including the monthly fee, constant when considering unlimited plans. To address this concern we compute the fees that maximize the ISP's revenues when offering unlimited plans, and find that 13.3% of users who subscribe under usage-based pricing plans do not subscribe in this counterfactual. Furthermore, the surplus of the remaining subscribers is lower. In total, the average per-subscriber monthly surplus is higher by $53.67 under usage-based pricing, and per-subscriber monthly revenues are lower by $51.65. Since traffic levels are virtually unchanged, we estimate no change in total welfare from usage-based pricing relative to this counterfactual.

Our second counterfactual evaluates welfare under the menu of usage-based plans offered in the data, relative to a situation where subscribers are presented with a single plan with unlimited usage and a one Gigabit per second (Gb/s) connection. This counterfactual relates to the first counterfactual in several ways. First, in the above we hold the set of plans fixed, while now we evaluate one particular new plan. Second, this plan has been offered as an alternative solution for rising traffic levels. Finally, this counterfactual allows us to highlight the importance of the preference for speed in generating value.

We choose a fixed fee of $100, which is within the range of unlimited plans with one Gb/s connections offered today. We find that approximately 48% of subscribers prefer the menu of usage-based pricing plans. Yet those subscribers who prefer the Gigabit plan derive substantially more surplus from the faster connection and unlimited usage. After accounting for the approximately 10% of subscribers who drop service when only offered the Gigabit plan, we find that total per-subscriber surplus generated from usage is $87.73 higher per month with the single high-speed plan. Yet the ISP's revenues only increase by $21.94. Thus, while total welfare increases by $87.73 per subscriber, the ISP only captures 25% of it, therefore there is a gap between private and social benefits from investment.

Proprietary estimates of the costs of building out a fiber-optic network capable of offering

[2] By total welfare we refer to the ISP profits and subscriber welfare. We cannot quantify the impact on the providers of content (e.g., NetFlix).

Gb/s speed are roughly $3,284 per customer (Kirjner and Parameswaran 2013). By our estimates, this requires a minimum of around 37 months to recover the capital expenditures from a social perspective, while the ISP recovers these costs only after about 150 months. This suggests these networks will become reality much later than is socially desirable without further subsidization. One caveat to these results is that we chose a fixed fee of $100 that may already reflect substantial subsidies. To examine the sensitivity of these results, and provide a measure of how much of a subsidy this price represents, we repeat the counterfactual allowing the ISP to choose the fixed fee. We find total welfare to be nearly unchanged, but a higher fee results in the ISP enjoying a larger share of the surplus.

The remainder of the paper is as follows. In the rest of this section we discuss some of the related literature. In Section 2 we discuss our data in greater detail and provide descriptive evidence to motivate assumptions made in the structural model. In Section 3 we discuss the model used to capture the intertemporal decisions of subscribers regarding consumption of content under usage-based billing. Section 4 presents our methodology for estimating the structural model, while Sections 5 and 6 discuss our results and counterfactual calculations, respectively. Section 7 concludes.

1.1 Related Literature

Our paper is related to past work that studies demand for broadband Internet access. Varian (2002) and Edell and Varaiya (2002) run experiments as part of the INDEX project, where users face varying prices for different allowances and speeds. They find most users value time spent online at less than 1 cent per minute. However, the experimental subjects were not given a choice of a compelling broadband option, and are strongly influenced by the limited online content and applications at the time of the study. Goolsbee and Klenow (2006) use Forrester Technographics Survey data on individuals' time spent on the Internet and earnings, jointly with an assumption that an hour spent on the Internet is an hour in forgone wages, to estimate the private benefit to subscribers of residential broadband. Lambrecht et al. (2007) use monthly consumption data from a German ISP to study the role of uncertainty in consumers' selection of usage-based plans. Several additional papers (Dutz et al. 2009; Rosston et al. 2010; Greenstein and McDevitt 2011) estimate the economic value of broadband internet, but without disaggregated decisions or usage-based pricing. Hitt and Tambe (2007) show that broadband adoption increases internet usage by 1,300 minutes per month, suggesting a strong preference for content that requires high bandwidth.

The modeling in this paper is related to two separate literatures. The first is a literature that focuses on estimating demand in dynamic settings. For example, Crawford and Shum (2005) develop an approach to estimate demand when consumers are actively learning about the quality of their match with a product (prescription drug). Hendel and Nevo (2006a) propose an approach to estimating demand when temporary price reductions induce stockpiling. Gowrisankaran and Rysman (2012) quantify the importance of accounting for dynamics in a market for a durable good with rapid price declines and constant quality improvements. Like our analysis, Yao et al. (2012) exploits intra-month (weekly) variation in the marginal or shadow price of usage under three-part tariffs to identify consumers' discount factors. The second is a literature studying incentives in labor contracts. Often, these contracts specify a nonlinear compensation structure based on performance during a fixed period of time, which makes the worker's decision regarding the optimal level of effort a dynamic one, in much the same way usage is under a three-part tariff. Notable studies in this literature include Copeland and Monnet (2009), Chung et al. (2010), and Misra and Nair (2011).

Another related literature questions whether consumers are forward-looking when making decisions. Notable studies in this literature include Aron-Dine et al. (2012), Chevalier and Goolsbee (2009), Grubb (2012), Grubb and Osborne (2012), and Hendel and Nevo (2006b). Like much of this literature we provide evidence that consumers respond to dynamic incentives.

2 Data

The data come from a North American ISP that sells access to the Internet over a cable broadband network. To obtain service, a subscriber buys a monthly service plan from a menu of choices. Features of a plan include maximum download and upload speeds, an access fee, usage allowance (if any), and overage price per GB of data (if any).[3] We say a subscriber who does not face a positive overage price is on an *unlimited* plan. The ISP offers only usage-based plans to new subscribers. All subscribers on unlimited plans are grandfathered in.

Subscribers are residential households who are uniquely identified by their cable modem. Usage in GBs is recorded for both the upstream (e.g., uploading a file to Dropbox) and downstream (e.g., streaming a movie from Netflix) directions, in the form of Internet Protocol Detail Record (IPDR) data. For billing purposes, and consequently our purposes, the direction of the traffic is ignored, and we examine the total traffic in either direction.

Each subscriber shares access to the Internet with a group of other subscribers, which raises the possibility that congestion externalities could affect network performance. Fortunately, our

[3] Subscribers are not on long-term contracts, only incurring a disconnection fee if service is canceled.

data come from a period when the network was overly-provisioned, virtually eliminating such externalities.[4] In modeling a subscriber's usage decision, we reasonably assume that their usage decision does not depend on concerns over congestion in the network.

2.1 Sample and Descriptive Statistics

The sample includes 54,801 subscribers in four different markets served by a single ISP. The residents of these four markets had per-capita income of $47,592 in 2011, relative to $45,222 for residents in all US metropolitan markets. Like many other US markets, the alternative to the service provided by this cable company is DSL through a telecom company whose fastest connection speed is less than half that of the cable company's fastest speed.

For each subscriber, we observe usage at the monthly level from May 1^{st}, 2011—May 31^{st}, 2012, and for 15-minute intervals during May 10^{th} to June 30^{th}, 2012. We also know the plan chosen by the subscriber.[5]

Table 1: *Usage, May 2011 and May 2012*

Percentile	**May 2011 (GB)**	**May 2012 (GB)**	**Growth (GB)**	**Growth (%)**
25	2.49	6.69	4.20	168.67
50	8.99	20.27	11.28	125.47
75	26.85	52.24	25.39	94.56
90	60.83	103.94	43.11	70.87
95	92.62	147.27	54.65	59.00
99	185.81	253.62	67.81	36.49
Mean	23.08	40.29	17.21	74.56

Note: Based on usage by 54,801 subscribers to a single ISP, in four markets during May 2011 and May 2012. Usage is computed by aggregating IPDR data, captured in 15-minute intervals, to the monthly level. Means and percentile statistics are at the subscriber level.

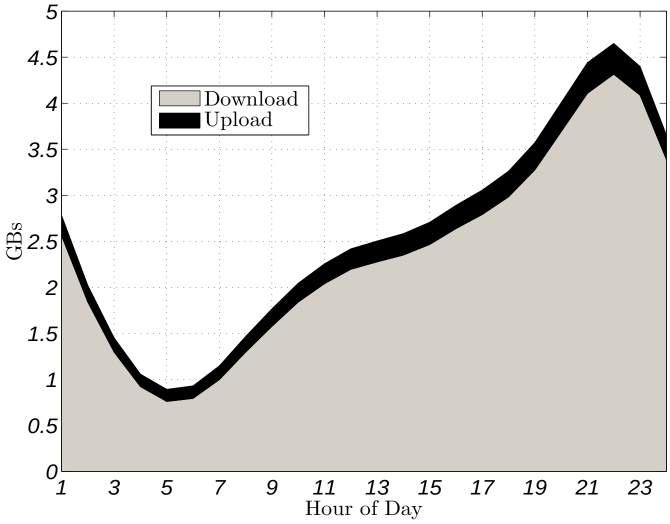
The rapid growth in usage can be seen in Table 1. The median subscriber's usage more than doubles, from 8.99 GBs in May 2011 to 20.27 GBs in May 2012. This increase of over 11 GB per month is equivalent to about eleven additional standard-definition movies or four additional high-definition movies. The average subscriber's usage increases more in absolute terms, from 23.08 GBs to 40.29 GBs, but less in percentage terms. Growth in absolute terms is monotonically

[4]IPDR data also capture "packets dropped," which measures the amount of internet traffic that does not reach its destination due to congestion. Our discussions with industry experts suggest that packets dropped in excess of 0.5% correspond to a degraded quality of service that would be noticeable to the subscriber. During May 10^{th} 2012—June 30^{th} 2012, no subscriber suffered more than 0.5% of packets dropped in any one-hour period.

[5]Unfortunately, we do not have information on bundling of other services. So we do not know if the subscriber is paying for the Internet service as part of a bundle.

higher for more intensive users, but in percentage terms is monotonically higher for less-intensive users.

Figure 1: *Distribution of Monthly Usage by Hour and Direction*



Note: This figure presents the total, upstream and downstream, traffic generated by the average user in each hour of the day for one complete billing cycle during May 10^{th}–June 30^{th}, 2012.

The more disaggregated data, which include one complete billing cycle for each subscriber during May and June of 2012, form the basis of our main analysis. Usage by these subscribers shows a standard cyclical pattern through the course of a day. Figure 1 plots the total, upstream and downstream, traffic generated by the average user in each hour of the day. Throughout the day, approximately 90% of the usage is in the downstream direction; in our modeling we will focus only on total traffic and not on the breakdown between upload and download. Peak usage occurs 10pm-11pm, when the average user consumes over 4.5 GBs each month. This is over four times the amount of traffic generated during 5am-6am. These patterns are qualitatively similar to *time* usage of dial-up and broadband Internet service reported by Rappoport et al. (2002).

Due to the sheer volume of data and the similarity in the directional and temporal patterns of usage across the day for the entire distribution of users, we aggregate usage to a daily level. This results in 1,274,550 subscriber-days, which is the unit of observation for our preliminary and structural analysis.

Table 2 reports summary statistics on monthly usage and plan characteristics for this billing cycle. These statistics highlight how allowances and overage prices curtail usage. An average subscriber to an unlimited plan pays $44.33 for a month of service, enjoys a maximum download speed of 6.40 Mb/s and uses just over 50 GB. In contrast, an average subscriber to a usage-

Table 2: *Descriptive Statistics of Subscriber Plan Choices and Usage, June 2012*

	Unlimited Plans	**Usage-Based Plans**
Number of Subscribers		
	12,316	42,485
Plan Characteristics		
Mean Access Fee ($)	44.33	74.20
Mean Download Speed (Mb/s)	6.40	14.68
Mean Allowance (GB)	∞	92.84
Mean Overage Price ($/GB)	0.0	3.28
Usage		
Mean (GB)	50.39	43.39
Mean (Access Fee $\leq$ $60) (GB)	48.94	20.40
Mean (Access Fee > $60) (GB)	88.59	69.36
Median (GB)	25.60	23.63
Median (Access Fee $\leq$ $60) (GB)	25.17	12.18
Median (Access Fee > $60) (GB)	42.12	52.04
Median Price per GB ($)	1.68	3.02

Note: These statistics reflect characteristics of plans chosen and usage by subscribers to a single ISP, in four markets during June 2012. Across plans, download speed is non-decreasing in the access fee and the overage price is non-increasing in the access fee. Usage is based upon Internet Protocol Detail Record (IPDR) data, captured in 15-minute intervals and aggregated to the monthly level. Means and medians are at the subscriber level. Among subscribers to unlimited plans, 11,868 choose a plan with an access fee less than $60. Among subscribers to usage-based plans, 22,529 choose a plan with an access fee less than $60.

based plan pays nearly $30 more per month to enjoy faster download speed (14.68 Mb/s), but uses under 44 GB. For both categories of subscribers, median usage is lower than mean usage. The median subscriber to an unlimited plan pays about $1.68 per GB, less than half what a usage-based subscriber pays.

2.2 Do Subscribers Choose the Optimal Plan?

Previous work has documented that people frequently make mistakes when confronted with complicated economic decisions (Thaler and Sunstein 2008). Moreover, economic research has highlighted mistakes by consumers facing non-linear pricing, similar to ours, in cell phone usage (Grubb and Osborne 2012) and health care (Abaluck and Gruber 2012; Handel 2013).[6] Grubb and Osborne (2012) find that $29 - 45\%$ of plan choices are suboptimal ex post. Abaluck and Gruber (2012) find that over 70% of elders choose plans that are not on the "efficient frontier"

[6]Goettler and Clay (2011) offer an alternative explanation, i.e., uncertainty and learning on the behalf of consumers, that can explain away apparent behavioral biases in some settings.

Table 3: *Descriptive Statistics, Usage-Based Plans*

	5/2011 - 5/2012	**6/2012**
Number of Subscribers	42,485	42,485
Mean Share of Allowance Used (%)	46.05	49.02
Subscribers Over Allowance (%)	8.62	9.45
Median Overage (GB)	14.31	17.03
Median Overage Charges ($)	44.98	51.19
Subscribers on Dominated Plan (%)	0.13	7.24

Note: These statistics reflect usage by subscribers to a single ISP, in four markets during May 2011-June 2012. Usage is based upon Internet Protocol Detail Record (IPDR) data, captured in 15-minute intervals and aggregated to the monthly level. Overage statistics reflect only those subscribers incurring positive overage charges. A subscriber is said to be on a dominated plan if there was an alternative to the chosen plan that would have yielded a lower per GB cost and with download speed at least as fast as the chosen plan. Means and medians are at the subscriber level.

of plans, in the sense that they could achieve better protection with less risk with another plan. Handel (2013) shows that consumers similarly fail to switch out of plans whose characteristics render them dominated.

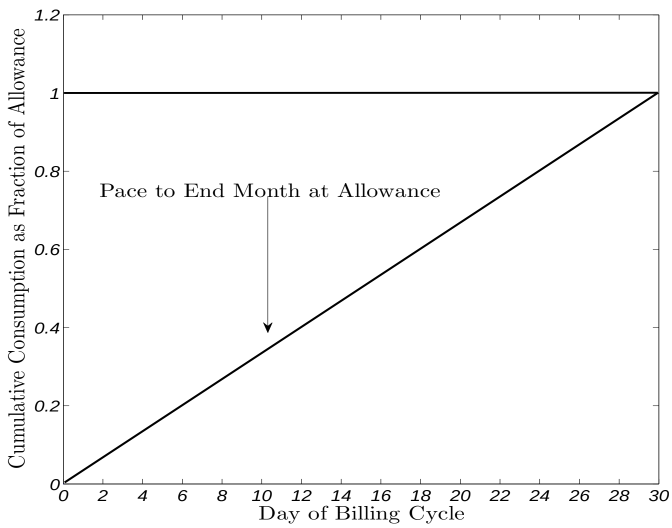
In Table 3, we report summary statistics for overage charges incurred and the frequency of different types of "mistakes" for subscribers on plans with usage-based pricing. During June 2012, about ten percent of subscribers on plans with usage-based pricing exceed their allowance. This is important, as our identification strategy relies on having enough subscribers solving a dynamic problem, i.e., some probability of incurring overage charges during the month. On average, subscribers use slightly less than half of their usage allowance, and of those that go over, the median amount over the allowance is 17.03 GBs.

In our data, it is less obvious that subscribers systemically make mistakes. One way to measure these mistakes, which is comparable to that documented by previous work for other industries, is to ask how many subscribers choose a dominated plan. In our context a dominated plan is one in which subscribers could have paid less and used service that is no slower. If one looks at the complete billing cycle (June 2012) in isolation, one would conclude that 7.24% of subscribers used a dominated plan. However, the frequency of this type of mistake goes down to 0.13% if we ask how many subscribers could have paid less and used service that is no slower during the 13 months from May 2011 to May 2012. To calculate this, we hold each subscriber's usage at its observed level. Given the infrequency of obvious mistakes in our data, our model assumes that consumers are rational in their plan choice and choose the optimal plan ex-ante.[7]

2.3 Are Subscribers Forward Looking?

We now provide evidence that subscribers are forward looking. This is interesting for two reasons. First, the evidence we provide adds to a growing literature demonstrating forward-looking behavior. Notable studies in this literature include Aron-Dine et al. (2012), Chevalier and Goolsbee (2009), Crawford and Shum (2005), Gowrisankaran and Rysman (2012), Hendel and Nevo (2006b), and Yao et al. (2012). Second, our identification relies on consumers responding to changes in the shadow price of usage over a billing cycle. It is therefore useful to know that consumers are indeed responding to this variation before proceeding to the structural model.

Figure 2: *State Space*



Note: This figure presents the set of states, i.e., a particular day in the billing cycle and cumulative usage up until that day as a proportion of the allowance, that determine a subscriber's marginal price for usage. Those with cumulative usage greater than the allowance face a constant marginal price equal to the overage price on their plan. Those with cumulative usage below their allowance face a shadow price equal to roughly the overage price times the probability they will ultimately exceed the cap.

Our data are from a provider that allows subscribers to carefully track their usage, by receiving text messages and emails at regular intervals after they exceed one-half of their allowance. Consumers may also log into the provider's web site at any time. We therefore have confidence that subscribers are aware of previous usage during the month.

The way in which past usage impacts current usage is through the marginal price of usage.

[7]In Section 3, we discuss how one might relax this assumption.

Figure 2 highlights this idea. Subscribers that have exceeded their allowance, i.e., past usage during the billing cycle as a proportion of the allowance is greater than one, face a marginal price equal to the overage price on their plan. Subscribers with past usage below their allowance effectively face a shadow price of usage equaling the discounted overage price times the probability they will ultimately exceed the allowance. The diagonal line in Figure 2 represents the usage rate that would result in a subscriber finishing the month having exactly exhausted their allowance. Thus the further above this line a subscriber is at a point in time, the greater the shadow price of usage.

If subscribers are forward looking, we expect certain patterns in usage throughout a billing cycle. The heaviest-volume subscribers that know they are likely to exceed the allowance, i.e., a high probability of exceeding their allowance by the end of the billing cycle, should behave as though the shadow price is equal to the overage price from the beginning of the billing cycle. For these subscribers, if they are forward looking, there should be little change in average usage throughout the billing cycle. Similarly, for subscribers with only a small probability of exceeding their allowance, i.e., those consistently below the diagonal line in Figure 2, behavior should not vary throughout the billing cycle. The only exception would be a small increase in usage towards the end of the billing cycle when the probability of exceeding the usage allowance approaches zero. For subscribers between these two extremes, usage should vary significantly depending on both the day in the billing cycle and a subscriber's cumulative usage up until that day.

To test whether consumers respond to the price variation introduced by past usage within a billing cycle, we estimate the following regression

$$ln(c_{jkt}) = \sum_{m=1}^{M=4}\sum_{n=1}^{N=5} \alpha_{nm} \mathbb{1}\left[pct_n \leq \frac{C_{jk(t-1)}}{\overline{C}_k} < pct_{n+1}\right] \mathbb{1}\left[day_m \leq t < day_{m+1}\right] \qquad (1)$$
$$+\mathbf{x}_t\psi + \mu_j + \epsilon_{jkt},$$

where the dependent variable, $ln(c_{jkt})$, is the natural logarithm of subscriber j's usage on day t, on plan k.

The ratio, $\frac{C_{jk(t-1)}}{\overline{C}_k}$, is the proportion of the usage allowance used up until day t, or the subscriber's total usage in the previous $(t-1)$ days of the billing cycle, $C_{jk(t-1)} = \sum_{\tau=1}^{t-1} c_{jk\tau}$, divided by the usage allowance on plan k, $\overline{C}_k$. The first set of indicators, $\mathbb{1}\left[pct_{n-1} \leq \left(\frac{C_{jk(t-1)}}{\overline{C}_k}\right) < pct_n\right]$, equals one when the proportion of a subscriber's usage allowance that has been used to date is in a particular range, such that $pct_1 = 0$, $pct_2 = 0.40$, $pct_3 = 0.60$, $pct_4 = 0.80$, $pct_5 = 1.00$, and $pct_6 = \infty$. The other set of indicators, $\mathbb{1}\left[day_{m-1} \leq t < day_m\right]$, equals one when the day is in a particular range, such that $day_1 = 10$, $day_2 = 15$, $day_3 = 20$, $day_4 = 25$, and $day_5 = 31$. We omit the interactions for the first ten days of the billing cycle, since there are so few subscribers

Table 4: *Forward-Looking Behavior, Within-Month Regression*

	$\mathbb{1}\,[10 \leq t < 15]$	$\mathbb{1}\,[15 \leq t < 20]$	$\mathbb{1}\,[20 \leq t < 25]$	$\mathbb{1}\,[25 \leq t < 31]$
$1\left[0 \leq \frac{C_{jk(t-1)}}{\overline{C}_k} < 0.40\right]$	-0.04**	-0.04**	0.03**	0.08**
	(0.01)	(0.01)	(0.01)	(0.01)
$1\left[0.40 \leq \frac{C_{jk(t-1)}}{\overline{C}_k} < 0.60\right]$	-0.02	-0.12**	-0.12**	-0.04**
	(0.02)	(0.01)	(0.01)	(0.01)
$1\left[0.60 \leq \frac{C_{jk(t-1)}}{\overline{C}_k} < 0.80\right]$	-0.07**	-0.12**	-0.20**	-0.16**
	(0.03)	(0.02)	(0.02)	(0.01)
$1\left[0.80 \leq \frac{C_{jk(t-1)}}{\overline{C}_k} < 1.00\right]$	-0.19**	-0.26**	-0.39**	-0.42**
	(0.05)	(0.03)	(0.02)	(0.02)
$1\left[1.00 \leq \frac{C_{jk(t-1)}}{\overline{C}_k}\right]$	-0.12**	-0.35**	-0.41**	-0.47**
	(0.05)	(0.03)	(0.02)	(0.02)
Adjusted R^2	0.46			

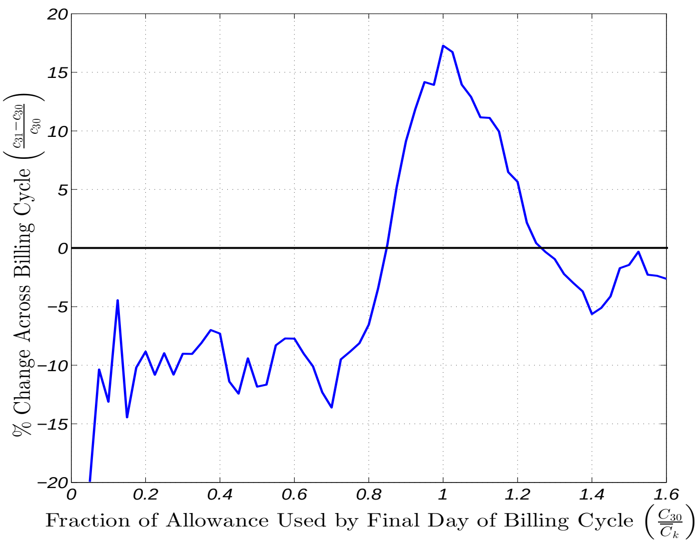
Note: This table presents OLS estimates of Equation (1) using 1,274,550 subscriber-day observations. The dependent variable is natural logarithm of daily usage. Each cell in the table gives the coefficient on the interaction between the indicators in the respective row and column. Controls include a constant, time trend, indicators for the day of the week, and subscriber fixed effects. Asterisks denote statistical significance: ** 1% level, * 5% level.

that have used a substantial proportion of their allowance by this time. Additional controls, $\mathbf{x}_t$, include dummy variables for the days of the week and a time trend to account for any organic growth in usage over the course of the billing cycle. The inclusion of subscriber fixed effects, μ_j, removes persistent forms of heterogeneity across subscribers.

The estimates of Equation (1) are reported in Table 4. Each cell reports the estimate for the coefficient on the interaction between the indicators in the respective row and column. At each point in the billing cycle, we find current usage to be responsive to past usage in a way that is consistent with forward-looking behavior. Subscribers who are near the allowance early in the billing cycle reduce usage substantially less than subscribers who near the allowance later in the billing cycle (i.e., coefficients are monotonically decreasing from left to right within rows four and five of Table 4). This is consistent with the heaviest-volume subscribers expecting to go through the allowance early in the billing cycle, and moderate-volume subscribers facing substantial uncertainty that is only resolved later in the billing cycle. That is, those nearing the allowance later in the billing cycle reduce usage proportionally more, relative to their own mean, than those nearing the allowance early in the billing cycle. For subscribers well below the allowance late in the billing cycle, we observe a small increase in usage, consistent with these subscribers becoming confident that they will not exceed the allowance.

Besides the within-month variation in price that subscribers encounter, there is variation as a new billing cycle begins. Specifically, subscribers experience a discrete change in the shadow price when their usage allowance is refreshed. A forward-looking subscriber near the allowance at the end of a billing cycle knows that the shadow price unambiguously decreases at the beginning of the next billing cycle. Conversely, a subscriber well below the allowance likely experiences an increase in the shadow price as the new billing cycle begins. Subscribers well over the allowance at the end of the billing cycle, who expect to go over the allowance again next month, should behave as though the price always equals the overage price and not respond at all.

Figure 3: *Across-Month Dynamics*



Note: This figure presents how the percentage change in usage from the last day of a billing cycle to the first day of the next varies with the proportion of the allowance consumed by a subscriber at the end of the billing cycle.

For most subscribers, we observe at least one day of usage beyond the full billing cycle used for the rest of our analysis, allowing for a test of whether subscribers respond to this across-month price variation. To do so, we first calculate the percentage change in usage from the final day of the billing cycle ($t = 30$) to the first day of the next billing cycle ($t = 31$) for each subscriber, $\frac{c_{jk(31)}-c_{jk(30)}}{c_{jk(30)}}$. We then calculate the mean percentage change for groups of subscribers that used various fractions of the allowance by the end of the month, $\frac{C_{30}}{\overline{C}_k}$. Figure 3 presents the results. Subscribers facing a price increase at the beginning of the next month consume relatively more at the end of the current month, while those expecting a price decrease consume relatively less. We observe little change in usage for those well above the allowance in the current month.

Collectively, our results provide support for the hypothesis that subscribers are forward looking. Consumers are responsive, in an economically meaningful way, to variation in the shadow price of usage both within and across billing cycles.

3 Model

We model the subscriber's problem in two stages. The subscriber first chooses a plan anticipating future demand for content, and then chooses usage given the chosen plan.

3.1 Utility

Subscribers derive utility from content and a numeraire good. To consume content, each subscriber chooses a plan, indexed by k. Each plan is characterized by the speed by which content is delivered over the internet, s_k, by the usage allowance, $\overline{C}_k$, by the fixed fee, F_k, and by the per-GB overage price, p_k. Specifically, F_k pays for all consumption up to $\overline{C}_k$, while overage consumption costs p_k per GB. For any plan, the number of days in the billing cycle is T.

Utility from content is additively separable over all days in the billing cycle.[8] Let consumption of content on day t of the billing cycle be c_t and the consumption of the numeraire good on day t be y_t. We specify a quasi-linear form, where a subscriber of type h on plan k has

$$u_h(c_t, y_t, \upsilon_t; k) = \upsilon_t \left(\frac{c_t^{1-\beta_h}}{1-\beta_h} \right) - c_t \left(\kappa_{1h} + \frac{\kappa_{2h}}{\ln(s_k)} \right) + y_t.$$

The first term captures the subscriber's gross utility from content and specifies it as random across days. The *curvature* of the isoelastic function is allowed to vary from log ($\beta_h \to 1$) to linear ($\beta_h = 0$). The time-varying unobservable, υ_t, is not known to the subscriber until period t. For type h, each υ_t is independently and identically distributed LN(μ_h, σ_h), truncated at point $\overline{\upsilon_h}$ to exclude the top 0.01% of the distribution. For simplicity, we denote type h's distribution of υ_t as G_h, and refer to μ_h and σ_h as the mean and standard deviation of the distribution.

The second term captures the subscriber's non-price cost of consuming online content. Marginal cost is constant, at $\kappa_{1h} + \frac{\kappa_{2h}}{\ln(s_k)}$. The parameter $\kappa_{1h} > 0$ captures the consumer's *opportunity cost of content*, notwithstanding wait time. The ratio $\frac{\kappa_{2h}}{\ln(s_k)}$, where $\kappa_{2h} > 0$ is the subscriber's *preference for speed*, captures the waiting cost of transferring content. This specification implies that the subscriber has a satiation point absent overage charges, which is important to explain why consumers on unlimited plans consume a finite amount of content.

The vector of parameters, $(\beta_h, \kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$, describes a subscriber of type h. Conditional

[8] 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 that has not previously had access to the internet.

on choosing plan k, this subscriber's problem is

$$\max_{\{c_1,\ldots,c_T\}} \quad \sum_{t=1}^{T} E\left[u_h(c_t, y_t, \upsilon_t; k)\right]$$

$$s.t. \quad F_k + p_k Max\{C_T - \overline{C}_k, 0\} + Y_T \leq I, \quad C_T = \sum_{j=1}^{T} c_t, \quad Y_T = \sum_{j=1}^{T} y_t.$$

We do not discount future utility since we are looking at daily decisions, over a finite and short horizon. Notice this formulation assumes that the subscriber is aware of cumulative consumption, C_{t-1}, on each day in the billing cycle. The only uncertainty involves the realizations of υ_t. We assume that I is large enough so that wealth does not constrain consumption of content, and henceforth ignore substitution of numeraire consumption across days. The subscriber's problem is then a stochastic finite-horizon dynamic program.

3.2 Optimal Consumption

Denote the unused allowance at the beginning of period t, for a subscriber on plan k, as $\overline{C}_{kt} \equiv Max\{\overline{C}_k - C_{t-1}, 0\}$. Similarly, denote period-t overage as $\mathcal{O}_{tk}(c_t) \equiv Max\{c_t - \overline{C}_{kt}, 0\}$.

In the terminal period (T) of a billing cycle, the efficiency condition for optimal consumption depends on whether it is optimal to exceed the allowance. Intuitively, for a subscriber well below the allowance (i.e., $\overline{C}_{kT}$ is high) and without a high draw of υ_T, it is optimal to consume content up to the point where $\frac{\partial u_h(c_t, y_t, \upsilon_t; k)}{\partial c_t} = 0$. If marginal utility at $c_t = \overline{C}_{kT}$ is positive but below p_k, then it is optimal to consume exactly the remaining allowance. For a subscriber who is already above the allowance (i.e., $\overline{C}_{kT} = 0$) or who draws a high υ_T, it is optimal to consume up to the point where $\frac{\partial u_h(c_t, y_t, \upsilon_t; k)}{\partial c_t} = p_k$.

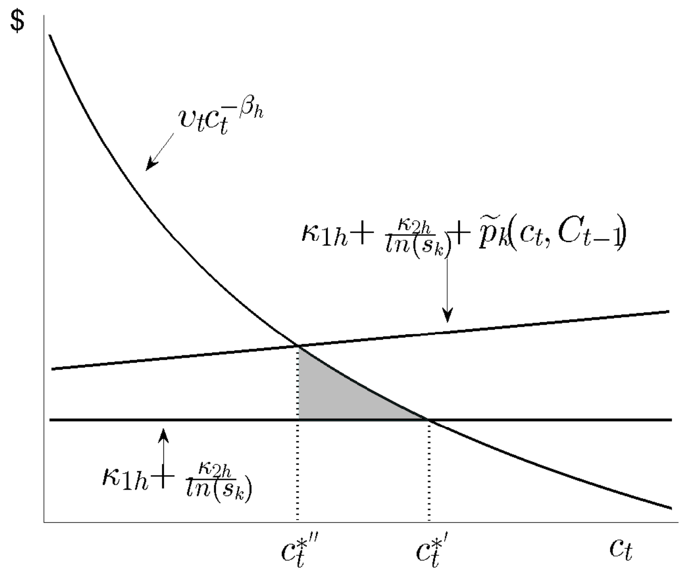
In the last period, there are no intertemporal tradeoffs. The subscriber solves a static utility maximization problem, given cumulative usage up until period T, C_{T-1}, and the realization of preference shock, υ_T. Denoting this optimal level of consumption by $c^*_{hkT}(C_{T-1}, \upsilon_T)$, the subscriber's utility in the terminal period is then

$$V_{hkT}(C_{T-1}, \upsilon_T) = \quad \upsilon_T \left(\frac{\left(c^*_{hkT}\right)^{1-\beta_h}}{1-\beta_h} \right) - c^*_{hkT} \left(\kappa_{1h} + \frac{\kappa_{2h}}{\ln(s_k)} \right) + y_t - p_k \mathcal{O}_{tk}(c^*_{hkT}).$$

For any other day in the billing period $t < T$, usage adds to cumulative consumption and affects the next period's state, so the optimal policy function for a subscriber incorporates this. Specifically, type h on plan k solves

$$c^*_{hkt}(C_{t-1}, \upsilon_t) = \underset{c_t}{argmax} \quad \left\{ \upsilon_t \left(\frac{c_t^{1-\beta_h}}{1-\beta_h} \right) - c_t \left(\kappa_{1h} + \frac{\kappa_{2h}}{\ln(s_k)} \right) + y_t - p_k \mathcal{O}_{tk}(c_t) + E\left[V_{hk(t+1)}(C_{t-1} + c_t) \right] \right\}.$$

Figure 4: *Optimal Consumption In Period t*



Note: The figure illustrates the optimal consumption when there are no overage charges, $c_t^{*'}$, and when the shadow price is positive, $c_t^{*''}$.

Define the *shadow price* of consumption

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

Then the consumer's optimal choice in period t satisfies

$$c_{hkt}^* = \left(\frac{\upsilon_t}{\kappa_{1h} + \frac{\kappa_{2h}}{ln(s_k)} + \widetilde{p}_k(c_{hkt}^*, C_{t-1})} \right)^{\frac{1}{\beta_h}} . \tag{2}$$

We illustrate this solution in Figure 4. When the subscriber is on an unlimited plan, or has cumulative consumption C_{t-1} such that the probability of exceeding the allowance is zero, then $\widetilde{p}_k(c_t, C_{t-1}) = 0$ and the optimal consumption is $c_t^{*'}$. When the subscriber is on a usage-based plan and the probability of exceeding the allowance is positive, then $\widetilde{p}_k(c_t, C_{t-1})$ is positive and the optimal choice is $c_t^{*''}$. If the probability the subscriber exceeds the allowance is strictly between 0 and 1, then $\widetilde{p}_k(c_t, C_{t-1})$ is upward-sloping up to the allowance because an additional unit of consumption today increases the probability that total monthly consumption will ultimately exceed $\overline{C_k}$. The shaded area shows the surplus lost under a move to usage-based pricing, assuming the subscriber does not switch plans.

Given the isoelastic specification of gross utility, we can view each subscriber as having a constant-elasticity inverse-demand function. Specifically, Equation (2) implies that a type with parameter β_h has demand elasticity equal to $-\frac{1}{\beta_h}$ with respect to changes in the *total disutility* of content, $\kappa_{1h} + \frac{\kappa_{2h}}{ln(s_k)} + \widetilde{p}_k(c_t, C_{t-1})$. The demand elasticity with respect to changes in $\widetilde{p}_k(c_t, C_{t-1})$ does not equal $-\frac{1}{\beta_h}$, however, because a percentage change in this price will not yield the same percentage change in the total disutility of content. Intuitively, a subscriber with curvature β_h will be less sensitive to changes in $\widetilde{p}_k(c_t, C_{t-1})$ than an elasticity of $-\frac{1}{\beta_h}$ implies.

The value functions are given by

$$V_{hkt}(C_{t-1}, \upsilon_t) = \upsilon_t \left(\frac{(c^*_{hkt})^{1-\beta_h}}{1-\beta_h} \right) - c^*_{hkt} \left(\kappa_{1h} + \frac{\kappa_{2h}}{\ln(s_k)} \right) + y_t - p_k \mathcal{O}_t \left(c^*_{hkt} \right) + E \left[V_{hk(t+1)}(C_{t-1} + c^*_{hkt}) \right]$$

for each ordered pair (C_{t-1}, υ_t). Then for all $t < T = 30$, the expected value function is

$$E \left[V_{hkt}(C_{t-1}) \right] = \int_0^{\overline{\upsilon_h}} V_{hkt}(C_{t-1}, \upsilon_t) dG_h(\upsilon_t).$$

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

$$E \left[c^*_{hkt}(C_{t-1}) \right] = \int_0^{\overline{\upsilon_h}} c^*_{hkt}(C_{t-1}, \upsilon_t) dG_h(\upsilon_t). \tag{3}$$

The solution to the dynamic program for each type (h) of subscriber implies a distribution for the time spent in particular states (t, C_{t-1}) over a billing cycle.

3.3 Optimal Plan Choice

Subscribers select plans before observing any utility shocks. Specifically, entering the first period with $C_0 = 0$, the subscriber selects plan $k \in \{1, ..., K\}$ to maximize expected utility. Alternatively, the subscriber may choose no plan at all, $k = 0$. Formally, the plan choice is given by

$$k^*_h = \underset{k \in \{0,1,...,K\}}{\arg\max} \left\{ E \left[V_{hk1}(0) \right] - F_k \right\}.$$

where the value $E \left[V_{h01}(0) \right]$ and the fixed access fee F_0 for $k = 0$, the outside option, are normalized to 0.

One could relax our assumption that consumers ex-ante choose the right plan by modeling the plan choice of each type of subscriber to maximize

$$\rho E \left[V_{hk0}(0) \right] + \epsilon_{hk} - F_k$$

where ϵ_{hk} are *iid* extreme value shocks and ρ is a parameter that scales the error shock and the expected utility from the plan. The parameter ρ can be considered the inverse of the variance of

ϵ. Given the infrequency of obvious ex-post mistakes in our data, we focus on the limiting case where the variance of ϵ_{hk} is equal to zero.

4 Estimation

We recover the joint distribution of the parameters using a method of moments approach similar to the two-step algorithms proposed by Ackerberg (2009), Bajari et al. (2007), and Fox et al. (2011). First, we solve the dynamic program for a wide variety of subscriber types, h. Second, we estimate the weight we should put on each of the types by matching the weighted average of optimal behavior, computed in the first stage, to the equivalent moments observed in the data. This yields an estimated distribution of types.

4.1 Step 1: Solving the Model

In the first step of the estimation algorithm we solve the dynamic problem for a large number of types (18,144, to be precise), where each type is defined by a particular value of the parameter vector $(\beta_h, \kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$. We solve the dynamic problem only once for each type and store the optimal policy so that we can form the moments needed in Step 2.

For a plan, k, and subscriber type, h, we solve the finite-horizon dynamic program described in the previous section recursively, starting at the end of each billing cycle ($t = T$). To do so, we discretize the C_t state to a grid of 2,000 points with spacing of size, Δc_k GBs, for each plan, k. The step size, Δc_k, is plan specific and non-decreasing in the plan's usage allowance, allowing for a denser state space on plans with lower usage allowances where usage is typically lower. The maximum consumption is set at five times the allowance for usage-based plans, and one Terabyte for unlimited plans, which is high enough to capture all usage in our data. Time is naturally discrete ($t = 1, 2, ..., 30$ over a billing cycle with $T = 30$ days) for our daily data. These discretizations leave υ_t as the only continuous state variable. Because the subscriber does not know υ_t prior to period t, we can integrate it out and the solution to the dynamic programming problem for each type of subscriber can be characterized by the expected value functions, $E\left[V_{hkt}(C_{t-1})\right]$, and policy functions, $E\left[c^*_{hkt}(C_{t-1})\right]$. To perform the numerical integration over the bounded support of υ_t, $[0, \overline{\upsilon}]$, we use adaptive Simpson quadrature.

Having solved the dynamic program for a subscriber of type h, we then generate the transition process for the state vector implied by the solution. The transition probabilities between the 60,000 possible states (2000*30) are implicitly defined by threshold values for υ_t. For example, consider a subscriber of type h on plan k, that has consumed C_{t-1} prior to period t. The threshold, $\upsilon_t(z)$, is defined as the value of υ_t that makes a subscriber indifferent between consuming z units of content of size Δc_k and $z+1$ units, such that the marginal utility (net of any overage charges) of an additional unit of consumption

$$u_h((z+1)\Delta c_k, y_t, \upsilon_t(z); k) - u_h(z\Delta c_k, y_t, \upsilon_t(z); k)$$

is equated to to the loss in the net present value of future utility

$$E[V_{hk(t+1)}(C_{t-1} + (z+1)\Delta c_k)] - E[V_{hk(t+1)}(C_{t-1} + z\Delta c_k)].$$

These thresholds, along with all subscribers' initial condition, $(C_0 = 0)$, define the transition process between states. For each subscriber type h and plan k, we characterize this transition process by the cdf of cumulative consumption that it generates,

$$\Gamma_{hkt}(C) = P\left(C_{t-1} < C\right), \tag{4}$$

the proportion of subscribers that have consumed less than C through period t of the billing cycle. Due to the discretized state space, $\Gamma_{hkt}(C)$ is a step function.

4.2 Step 2: Estimation

The second step of our estimation approach matches empirical moments we recover from the data to those predicted by our model by choosing weights for each subscriber type.

4.2.1 Objective Function

Our estimates of these weights are chosen to satisfy

$$\widehat{\theta} = \arg\min_{\theta} \mathbf{m}_k(\theta)'\widehat{\mathbf{V}}^{-1}\mathbf{m}_k(\theta),$$

subject to

$$\sum_{h=1}^{H_k} \theta_h = 1 \quad \text{and} \quad \theta_h \geq 0 \quad \forall h.$$

The plan-specific vector $\mathbf{m}_k(\theta)$ is given by,

$$\mathbf{m}_k(\theta) = \widehat{\mathbf{m}}_k^{dat} - \mathbf{m}_k^{mod}\theta \ ,$$

where $\widehat{\mathbf{m}}_k^{dat}$ is the vector of moments recovered from the data, and $\mathbf{m}_k^{mod}\theta$ is a weighted average of the equivalent type-specific moments predicted by the model. The $\mathbf{m}_k(\theta)$ vector has length equal to the dimension of the state space (60,000) times the number of distinct moments that are matched at each state (2), a total of 120,000 moments to be matched for each plan. The

$\mathbf{m}_k^{mod}$ matrix has H_k columns, where each column contains the moments predicted by the model for each of the H_k types that optimally select plan k. We define the moments we use and discuss how they are recovered from the data in Sections 4.2.2 and 4.2.3, respectively. The weighting matrix, $\widehat{\mathbf{V}}^{-1}$, is the variance covariance matrix of $\widehat{\mathbf{m}}_k^{dat}$, ensuring that more variable moments receive less weight.

After estimating the weights associated with each type that selects a plan, we appropriately normalize the weights to reflect the number of subscribers choosing each plan to get the joint distribution of types across all plans. The weights for each plan are estimated separately to appropriately deal with grandfathered unlimited plans, which may contain a type of subscriber that is also on a usage-based plan.

As pointed out by Bajari et al. (2007) and Fox et al. (2011), least squares minimization subject to linear constraints, and over a bounded support, is a well-defined convex optimization problem. Even though the optimization is over a potentially large number of weights, it is quick and easy to program in standard software as long as the moments are linear in the weights. This approach, as with any linear regression, requires that the type-specific matrix of moments predicted by the model for each plan ($\mathbf{m}_k^{mod}$) be of full rank. If types are too similar in their behavior, collinearity issues arise, and it is not possible to separately identify the weights associated with each type.

We therefore define the grid of types we consider in the following way. First, we rule out any types whose monthly consumption on an unlimited plan, with maximum speed, averages above one Terabyte. In the data, no subscriber uses more than one Terabyte in a single month. Second, we experiment with spacing of the grid, to ensure that the matrices are of full rank (to numerical standards). This process results in a grid of 18,144 types.

4.2.2 Choice of Moments and Identification

In choosing which moments to match we focus on two considerations: identification and computational ease. We discuss identification considerations below. Computation is significantly simpler if the moments are linear in the weights. To balance these considerations we choose the following two sets of moments.

First, we use the mean usage at each state

$$\sum_{h=1}^{H} E\left[c_{hkt}^{*}(C_{t-1})\right]\gamma_{hkt}(C_{t-1})\theta_h,$$

where $E\left[c_{hkt}^{*}(C_{t-1})\right]$ is the mean usage of type h in time t under plan k and past usage of C_{t-1}, and $\gamma_{hkt}(C_{t-1})$ is the probability that this type reaches the state. Note that the average is taken across all types on the plan, not just those that arrive at the state with positive probability. The reason we focus on the average across all users, and not just those that arrive in this state with positive probability, is a computational one. If we focus only on those that arrive with positive probability, then the moment will be non-linear in the parameters. The moment we use is easily computed from the data, and most importantly for us is linear in the θ_h weights.

The second set of moments is the mass of subscribers at a particular state

$$\sum_{h=1}^{H} \gamma_{hkt}(C_{t-1})\theta_h.$$

Like the first set of moments these moments are easy to compute from the data and are linear in the weights.

The estimation procedure recovers the weights of each type by asking which mixture of types results in predicted behavior that best matches the data. The "error term" is sampling error in the empirical moments. Thus, since the objective function is linear in the weights, the intuition for how the weights are identified is similar to that of a linear regression. Simply, the weights are identified as long as the behavior predicted by different types is not collinear. Thus, the key to identification is to understand how each parameter impacts the predicted behavior used to match the data.

To gain intuition consider the following example. Recalling Equation (2), fix $\beta_h = 1$ and consider the (average) consumption of a type h subscriber on plan k when the shadow price is constant (either because the allowance has been exceeded or because the probability of exceeding the allowance is zero) and equal to p. In this case average consumption is given by

$$E\left[c^*_{hkt}\right] = \frac{E(\upsilon_t)}{\kappa_{1h} + \frac{\kappa_{2h}}{ln(s_k)} + p}.$$

Note that $E(\upsilon_t)$ depends on both μ_h and σ_h. Thus with variation in speed and overage charges, different values of the $(\kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$ imply different behaviors across plans and states even when prices are constant.

In reality, focusing only on states where the price is constant is unlikely to be very informative, since we observe a limited number of speed options and overage charges. Our analysis, therefore, extends the same idea to a larger set of states. The moments we form from our data reveal both the proportion of subscribers that reach different states and how the consumption patterns of these subscribers change with variation in the prices across these states. For each combination of parameters, or a subscriber type, our model then predicts the probability of a type reaching different states and their behavior at those states. Identification of a mixture of types to best match the data (i.e, a unique minimum to the constrained least-squares objective function) is then ensured as long as one considers enough states such that every pair of types behaves differently at one or more states reached with positive probability.

While the nonlinearity of the parameters makes defining the role of each in influencing behavior, and subsequently differentiating how behaviors will differ for various types at different states, difficult, each parameter has a key role in determining particular types of behaviors. For example, the preference for speed, κ_{2h}, is critical for determining plan selection. Similarly, the curvature parameter, β_h, is important in determining how subscribers respond to price variation across the state space on a given plan. Similar arguments can be made for the other parameters.

4.2.3 Recovering Empirical Moments

The richness of the data, along with the low dimensionality of the state space, (C_{t-1}, t), allows a flexible approach for recovering moments from the data to match with the model.

Distribution of Cumulative Consumption To recover the cumulative distribution of C_{t-1} for each day t and plan k, we use a smooth version of a simple Kaplan-Meier estimator,

$$\widehat{\Gamma}_{kt}(C) = \frac{1}{N_k} \sum_{i=1}^{N_k} 1\left[C_{i(t-1)} < C\right].$$

We estimate these moments for each k and t, considering values of C such that $\widehat{\Gamma}_{kt}(C) \in [0.0001, 0.9999]$, ensuring that we fit the tails of the usage distribution. We use a normal kernel with an adaptive bandwidth to smooth the empirical cdf.

Mean Usage We recover the moments of usage at each state by estimating a smooth surface using a nearest-neighbor approach. Consider a point in the state space, (C_{t-1}, t). A neighbor is an observation in the data for which the subscriber is t days into the billing cycle and cumulative consumption up until day t is within five percent of C_{t-1}. Denote the number of neighbors by $N_{kt}(C_{t-1})$. Then, we estimate the conditional (on reaching the state) mean at (C_{t-1}, t) using

$$\widehat{E}\left[c^*_{kt}(C_{t-1})\right] = \frac{1}{N_{kt}(C_{t-1})} \sum_{i=1}^{N_{kt}(C_{t-1})} c_i,$$

where $i \in \{1....N_{kt}(C_{t-1}\})$ indexes the set of nearest neighbors. If $N_{kt}(C_{t-1}) > 500$, we use those 500 neighbors nearest to C_{t-1}. Note that this gives us the average usage conditional on a subscriber arriving at the state. To recover the unconditional mean,[9] we multiply $\widehat{E}\left[c^*_{kt}(C_{t-1})\right]$

[9] That is, the average of all subscribers, not just those that reach the state with positive probability

Table 5: *Descriptive Statistics, Type Distribution*

	Min	Max	Median	Mean	S.D.
mean of shocks (μ)	-0.25	1.50	1.25	1.06	0.83
s.d. of shocks (σ)	0.10	0.90	0.90	0.77	0.61
opp cost of content (κ_1)	0.50	10.50	6.50	5.68	4.37
pref for speed (κ_2)	0.50	10.50	2.50	5.13	4.41
curvature (β)	0.20	0.80	0.30	0.41	0.35

Note: These statistics reflect the estimated distribution of types after removing those types with weights $\theta_h <$ 0.0001 and renormalizing the weights of the remaining 62 types.

by the probability of observing a subscriber at state (C_{t-1}, t), recovered from the estimated cdf of cumulative consumption.

We estimate both moments at the same set of state space points used when numerically solving the dynamic programming problem for each subscriber type. This results in 120,000 moments for each plan. To compute the variance-covariance matrix, $\widehat{\mathbf{V}}_k$, of the resulting vector of moments, $\widehat{\mathbf{m}}_k^{dat}$, we draw on the literature on resampling methods with dependent data (e.g. Lahiri 2003). The dependence in the data comes from its panel nature, as we observe individuals making daily decisions on consumption over a full billing cycle. We repeatedly estimate $\widehat{\mathbf{m}}_k^{dat}$, leaving out different groups, or blocks, of subscribers. Specifically, we choose 1,000 randomly sampled groups of 5,000 subscribers and re-estimate the moments omitting a different group of subscribers each time.

5 Results

We estimate a weight greater than 0.01% ($\theta_h > 0.0001$) for 62 types. The most common type accounts for 43% of the total mass, the top five types account for 73%, the top 10 for 83% and the top 30 for 96%. No plan has more than 20 types receiving positive weights, while the average number of types across plans is only 7.75.[10]

The statistics in Table 5 highlight the non-normality of the type distribution. The distributions of the mean and the standard deviation of the random shocks, μ and σ, respectively, and the opportunity cost of content, κ_1, are left-skewed, while the distributions of the preference for speed, κ_2 and the utility curvature parameter, β, are right-skewed. The most common type is $(\beta_h = 0.30, \kappa_{1h} = 8.50, \kappa_{2h} = 2.50, \mu_h = 1.25, \sigma_h = 0.90)$. For this type, with average speed on

[10]Expanding the grid of types to allow for two additional values of each parameter, one above and below the current upper and lower limits, respectively, results in estimates that assign no weight greater than 0.01% to any of the additional types.

usage-based plans of 14.68 Mb/s, the intrinsic cost of consuming content, given by $\kappa_1 + \frac{\kappa_2}{ln(s_k)}$, is about \$9.43/GB, and average daily consumption (in absence of overages) would be about 1.7 GB and gross willingness-to-pay would be about \$208.

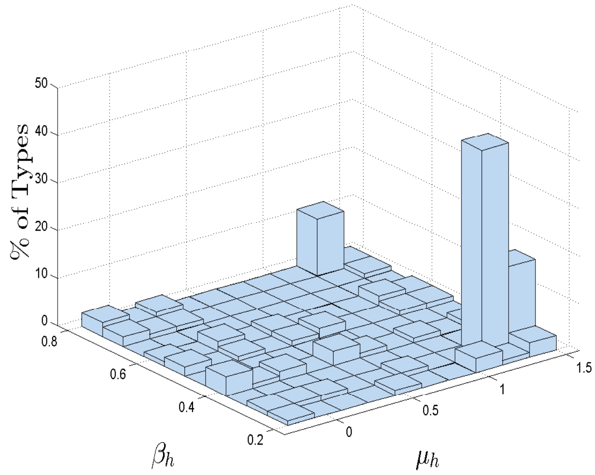
Overall, our model fits the data quite well. For all plans, the correlation between the empirical moments, $\widehat{\mathbf{m}}_k^{dat}$, and the fitted moments, $\widehat{\mathbf{m}}_k^{mod}\widehat{\theta}$, is above 0.99. Thus the model successfully replicates the average usage and density of subscribers at each state observed in the data, across a very diverse set of plans ranging from nearly linear tariffs to unlimited usage. The structural model also does well fitting patterns from the data that weren't explicitly matched during estimation. Consider the estimates from Figure 3, which relates how subscriber behavior changes as a new billing cycle begins to how much the subscriber consumed in the previous cycle. Using the structural estimates, we estimate that those subscribers consuming less than 80% of their allowance by the end of the billing cycle will decrease usage on the first day of the new cycle by 8.22%. For those users between 80% and 120% of their allowance, we estimate a 9.86% increase, while those over 120% increase usage by 0.21%. Each estimate is similar to those in Figure 3.

To demonstrate the importance of allowing for many types, we explore how the fit varies when the number of types is restricted. One straightforward way is to choose a small number of types with the largest weights on each plan (from the unrestricted estimates), re-optimize the weights over just these types, and evaluate the fit.[11] If we re-optimize over the weights of the two (three) most common types on each plan, we calculate the correlation between the empirical and fitted moments to be 0.86 (0.91). If we restrict the number of types to just the most common type of on each plan, which does not require re-optimization, the correlation is only 0.74. This deterioration in fit demonstrates the importance of allowing for a flexible approach in our application, making a single type or even a few types inadequate to replicate closely the behaviors observed on each plan.

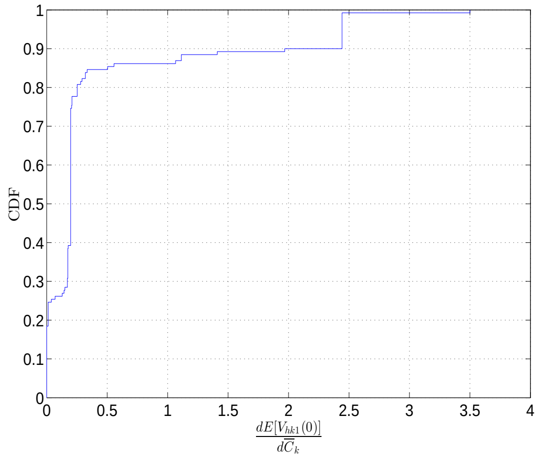
Figure 5(a) presents the joint distribution of the utility curvature, β_h, and the mean of the distribution of random shocks, μ_h. This distribution is highly irregular and non-normal. For the highest-volume subscribers (high μ_h), there is substantial variation in the elasticity of demand. In fact, for high-μ_h subscribers the distribution of β_h is clearly multi-peaked (unconditional on values of other parameters). The majority of high-volume subscribers have highly elastic demand, a value of β_h less than or equal to 0.3, including the most common type of subscriber. Most of the remainder of the high-μ_h subscribers have less elastic demand, or a value of β_h greater than or equal to 0.7.

[11] This calculation does not account for the possibility that the types with the largest weights from the unrestricted optimization may not provide the best fit among all combinations of types when the number of types is restricted.

Figure 5: *Joint Distribution of Utility Curvature (β_h) and the Mean of Shocks (μ_h); and Distribution of Willingness to Pay to Increase Usage Allowance by 1 GB*



(a) *Joint Distribution of β_{2h} and μ_h*



(b) *Value of Increasing Usage Allowance by 1 GB*

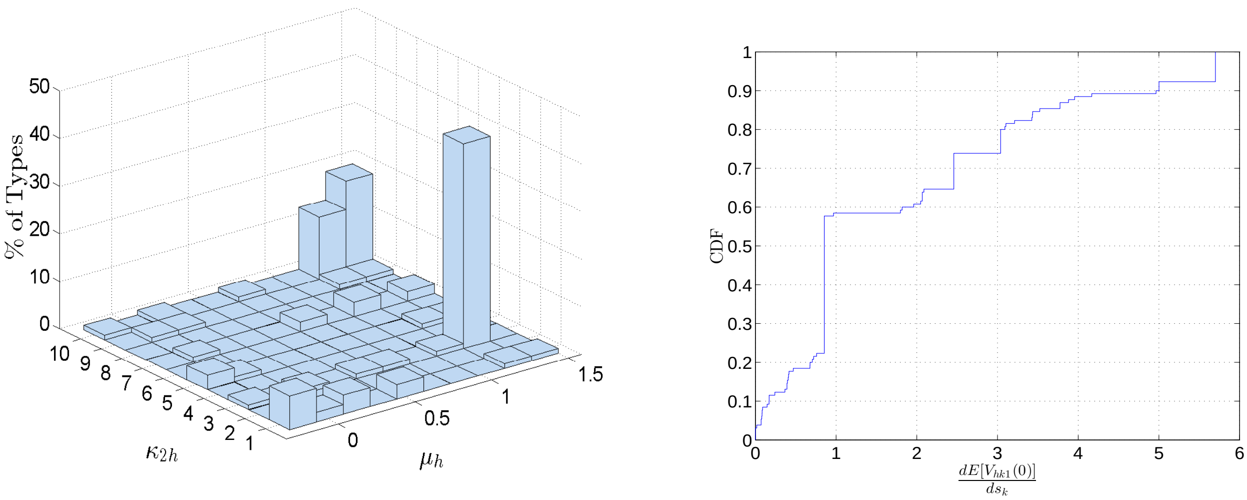
Note: Figure 5(a) shows the estimated weights for the joint distribution of the utility curvature (β_h) and mean of shocks (μ_h). Figure 5(b) shows the distribution of willingness to pay to increase usage allowance by 1 GB.

To better visualize what the distribution in Figure 5(a) implies about demand, Figure 5(b) shows the willingness to pay to increase the usage allowance by one GB on the first day of the billing cycle $\left(\frac{dE[V_{hk1}(0)]}{d\overline{C}_k}\right)$.[12] We note that approximately eighty percent of subscribers have a positive probability of incurring overage charges and would be willing to pay to increase their allowance if given the opportunity. The average (median) willingness to pay for a one GB increase is \$0.45 (\$0.23), and the distribution is left-skewed with a small number of subscribers who are willing to pay substantial amounts.

Figure 6(a) presents the joint distribution of the preference for speed, κ_{2h}, and the mean of the distribution of random shocks, μ_h. For the highest-volume subscribers (high μ_h), the marginal value placed on connection speeds has a wide range of values. As in Figure 5(a), the conditional distribution of κ_{2h} for high-μ_h values is clearly non-normal and multi-peaked. A relatively small group of individuals places high value on increased connection speeds (high κ_{2h}), but the majority of high-μ_h subscribers have a relatively low preference for speed. Despite this significant variation in the preference for speed, the overall value placed by subscribers on improving connection speeds is substantial. Figure 6(b) presents the distribution of $\frac{dE[V_{hk1}(0)]}{ds_k}$,

[12] For those subscriber types on unlimited grandfathered plans, we identify the optimal usage-based pricing plan currently offered by the ISP and perform the calculation. A total of 6.24% of subscribers choose not to subscribe to any usage-based pricing plan, and so they are omitted from Figure 5(b).

Figure 6: *Joint Distribution of Preference for Speed (κ_{2h}) and the Mean of Shocks (μ_h); and Distribution of Willingness to Pay to Increase Speed by 1 Mb/s*



(a) *Joint Distribution of κ_{2h} and μ_h*

(b) *Value of Increasing Speed by 1 Mb/s*

Note: Figure 5(a) shows the estimated weights for the joint distribution of the preference for speed (κ_{2h}) and mean of shocks (μ_h). Figure 5(b) shows the distribution of willingness to pay to increase speed by 1 Mb/s.

which is the rate of change in expected utility over the billing cycle from an increase in connection speeds (measured in Mb/s).[13] The value placed on improving speed by one Mb/s ranges from nearly zero to \$5.86, the average is \$1.76 and the median is \$0.87.

To further visualize what our estimates imply for demand, we consider subscriber behavior under a linear tariff. Suppose the ISP eliminates access fees and instead meters service at price p per GB. Further, for simplicity, suppose the ISP offers just one download speed s. Because there is no fixed fee, every subscriber type consumes something under this plan. And since this is a linear tariff, there are no dynamics.

Therefore, conditional on υ_t, a subscriber of type h chooses consumption according to Equation (2), with $s_k = s$ and $\widetilde{p}_k(c_t, C_{t-1}) = p$. Taking expectations over G_h for each type, and averaging across subscriber types, expected daily demand for content is then

$$D(p) = \sum_{h=1}^{H} \widehat{\theta}_h \int_0^{\overline{\upsilon}_h} \left(\frac{\upsilon}{\kappa_{1h} + \frac{\kappa_{2h}}{ln(s)} + p} \right)^{\frac{1}{\beta_h}} dG_h(\upsilon).$$

We estimate expected demand for three different speeds: (1) 2 Mb/s, a very slow speed by the standards of most contemporary broadband Internet service providers; (2) 14.68 Mb/s, the

[13]The same 6.24% of subscribers omitted from Figure 5(b) are also omitted here (see Footnote 11).

Table 6: *Expected Daily Usage Under a Linear Tariff*

	Speed=2 Mb/s		Speed=14.68 Mb/s		Speed=1,024 Mb/s	
Price	**Mean Usage**	**Elasticity**	**Mean Usage**	**Elasticity**	**Mean Usage**	**Elasticity**
0.00	0.56	0.00	2.38	0.00	5.14	0.00
1.00	0.38	-0.29	1.39	-0.39	2.70	-0.48
2.00	0.29	-0.47	0.98	-0.64	1.75	-0.79
4.00	0.19	-0.75	0.56	-1.03	0.87	-1.23
8.00	0.10	-1.12	0.23	-1.47	0.32	-1.68
16.00	0.04	-1.43	0.07	-1.80	0.09	-1.96

Note: This table presents the expected daily usage and elasticity averaged across all subscriber types when facing a linear tariff.

average speed for subscribers in our data; and (3) 1,024 Mb/s, the highest speed currently offered in North America. These are shown in Table 6.

For average speed, subscribers facing a zero price would consume an average of 2.38 GB per day, or roughly 71 GB per month. The average subscriber would reduce usage to about 0.5 GB per day when facing a price of $4 per GB. Average willingness to pay with a price of zero is about $9.36 per day, roughly $280 per month. These numbers suggest that, on unlimited plans, subscribers with average download speeds reap significant surpluses. Table 6 also shows price elasticities of expected demand. Elasticity is -0.4 at a price of $1, while at a price of $16, elasticity is about -1.8. At a price of about $4 per GB, the ISP earns maximal (single-price) revenue per day of about $2.22 per subscriber, a small portion of maximal willingness to pay.

For a speed of 2 Mb/s, expected usage is more than 75% lower, at just 0.56 GB per day, when the price is zero. Subscribers' willingness-to-pay falls by comparably less, to about $4.36. Demand is also less elastic. Intuitively, waiting costs form a much greater part of the subscriber's overall costs from consuming content, so price has less effect.

For a speed of 1,024 Mb/s, expected usage with a zero price is 5.14 GBs per day, just over twice as high as with average speed. Demand is more price elastic, as waiting costs are much lower. These estimates demonstrate how much current usage is limited by offered connection speeds, and how the volume of Internet traffic is likely to increase following investment in high-speed next-generation networks (e.g. fiber-optic or DOCSIS 3.1). We explore this further in Section 6.

6 Counterfactuals

We now conduct several counterfactual exercises to study the welfare implications of different solutions proposed to managing the rapid growth in Internet traffic.

6.1 Welfare Implications of Usage-Based Pricing

We begin by considering the impact of usage-based pricing (UBP) on subscriber welfare, and the ISP's costs and revenues. To do this we compare behavior under UBP to the case when consumers face the same set of plans except that allowances are unlimited, or equivalently there are no overage charges (i.e., $p_k = 0\ \forall k$). For the UBP case, we eliminate all grandfathered plans but otherwise keep plan features the same as in the data.

In this case subscriber welfare (weakly) decreases with UBP since the cost of Internet access (weakly) increases. The effect on total welfare is less clear and in principle may rise or fall under UBP. Total welfare depends on subscriber welfare as well as the revenue earned by the ISP and the costs of running the network, which are directly linked to usage. Our model of subscriber choice gives us direct estimates of revenues as well as estimates of total usage. Interestingly, as we discuss below, some consumers actually increase their usage under UBP. Below, we start by abstracting from network costs and discuss total surplus as the sum of subscriber welfare and revenue. At the end of this subsection, we touch on network costs.

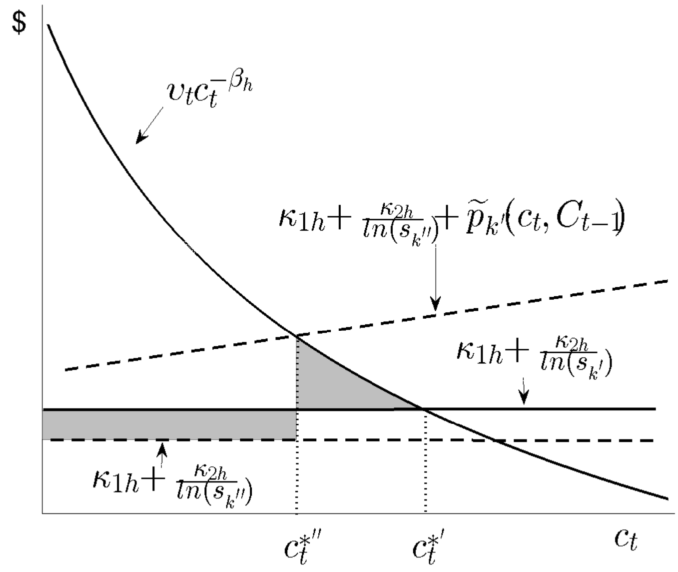
Subscribers fall into three groups: (1) those who choose the same plan under the unlimited and UBP menus—about 20% of the population according to our estimates; (2) those who choose different plans under the two menus—about 74% of the population; and (3) those who purchase service under the unlimited menu but do not purchase service under UBP—about 1% of the population.

The consumers in category (1) generate nearly the same revenue (absent any overage charges), since their plan choice is the same in both scenarios. Furthermore, usage decreases as their price is weakly higher under UBP. Thus, total welfare generated from this group is lower. Similarly, the total welfare from consumers in category (3) decreases under UBP as a welfare-enhancing trade is eliminated.

The most interesting welfare effects are for those consumers in category (2). Within this category, subscribers may switch to plans with lower or higher speed. Subscribers switching to a lower-speed plan are simple to analyze. They lower their usage, endure higher waiting costs and generate less surplus.

Subscribers who switch to a higher-speed plan are more complicated to analyze, because

Figure 7: *Usage and Surplus Under Unlimited and Usage-Based Plans*



Note: The figure illustrates one potential outcome for usage and surplus for a subscriber switching from unlimited service with speed of $s_{k'}$ to a faster usage-based plan with speed of $s_{k''}$.

they may either increase or decrease usage under UBP. It might seem surprising that consumers might increase consumption when the cost of Internet access (weakly) increases under UBP, but the intuition is as follows. Consider two plans with speeds $s_{k'}$ and $s_{k''}$ where $s_{k'} < s_{k''}$, and consider a subscriber who chooses speed $s_{k'}$ when plans are unlimited. Under UBP, the plan with speed $s_{k''}$ carries a higher allowance than the plan with speed $s_{k'}$, so the subscriber might find it beneficial to switch to the faster plan. Once on the faster plan, per-GB waiting costs are lower by $\frac{\kappa_{2h}}{ln(s_{k'})} - \frac{\kappa_{2h}}{ln(s_{k''})}$. When deciding how much to consume, if this difference in waiting costs exceeds the shadow price under UBP, $\widetilde{p}_k(c_t, C_{t-1})$, then total disutility per-GB of content is lower under UBP, and the subscriber consumes more content. So while total cost of consuming Internet content goes up under UBP the marginal cost might decrease.

For those subscribers who switch to a plan with higher speed, computing the change in welfare is even further complicated. The utility from consumption consists of two parts: the utility from consuming content and the disutility from the per-GB waiting cost, given by $\kappa_{1h} + \frac{\kappa_{2h}}{ln(s_{k''})}$. With a faster plan the cost term is lower, and therefore utility is higher. If the subscriber consumes more content, as we discussed in the previous paragraph, then utility increases. Once we account for the higher monetary cost the consumer is worse off, but the monetary cost is a transfer to the ISP and therefore total surplus generated by this consumer is higher.

If a subscriber consumes less on the higher-speed plan, then the two terms have opposing

Table 7: *Usage-Based Pricing Counterfactual, Usage and Welfare*

	Same Plan		**Switch Plan**		**Only Unlimited**	
	Unlimited	**UBP**	**Unlimited**	**UBP**	**Unlimited**	**UBP**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Percent of Types (%)	20.23		73.53		1.31	0
Usage and Surplus						
Speed (Mb/s)	16.23	16.23	10.43	13.05	10	–
Usage (GBs)	77.76	62.57	36.47	34.74	23.74	–
Consumer Surplus ($)	189.25	177.40	153.96	138.13	4.84	–
Revenue ($)	81.07	81.14	43.25	64.48	39.99	–
Total Surplus ($)	270.32	258.54	197.21	202.61	44.83	–
Δ in Total Surplus ($)	-11.78		5.40		44.83	–

	Same Plan		**Switch Plan**		**Only Unlimited**	
	$\uparrow$ Usage	**$\downarrow$ Usage**	**$\uparrow$ Usage**	**$\downarrow$ Usage**	**Unlimited**	**UBP**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Percent of Types (%)	0	20.23	65.63	7.90	1.31	0
Mean Type						
mean of shocks ($\bar{\mu}_h$)	–	0.91	1.28	0.49	0.24	–
s.d. of shocks ($\bar{\sigma}_h$)	–	0.75	0.82	0.74	0.78	–
opp cost of content ($\bar{\kappa}_{1h}$)	–	4.88	6.67	2.54	2.35	–
pref for speed ($\bar{\kappa}_{2h}$)	–	7.04	5.12	3.56	0.70	–
curvature ($\bar{\beta}_h$)	–	0.44	0.41	0.38	0.41	–

Note: These statistics reflect a comparison between two counterfactual scenarios. In both scenarios the consumers face the same set of plans, with the same characteristics, as we see in the data. However, in the unlimited scenario, all plans have unlimited allowances and no overage charges. Although the ISP in our data does offer unlimited plans, the characteristics of those plans are different from the hypothetical characteristics of unlimited plans considered in this counterfactual. Except for the percentages, all numbers are per-subscriber averages.

welfare effects. The per-GB waiting cost is lower, which increases utility, but usage is lower, which lowers utility. The effect on surplus is ambiguous. We illustrate this point in Figure 7. Consumption under the unlimited plan, with speed $s_{k'}$, is denoted by $c_t^{*'}$, while consumption under UBP with a plan with higher speed, $s_{k''}$ is given by $c_t^{*''}$. The reduction in the waiting cost is given by the gray rectangle. The welfare loss due to lower usage, on the UBP plan versus the unlimited plan, is shown in the gray triangle. If the triangular area exceeds the rectangular area, then total surplus (due to the subscriber's usage) falls.

Table 7 summarizes the results for consumers in categories (1)-(3). The impact of UBP for consumers in category (1), those who keep the same plan, is summarized in Columns 1 and 2 of the top portion of Table 7. Average consumer surplus is lower by \$11.85 per month, while average monthly usage is lower by about 15.2 GBs. Average per-subscriber monthly total surplus generated by these consumers' usage is lower by \$11.78 because the ISP's revenues are essentially unchanged. The bottom panel of Table 7 describes the mean type of subscriber in this category. Since subscribers in this category do not switch plans, they all decrease usage. The means of the parameters largely reflect a selection effect, as those subscribers who do not switch plans when UBP is introduced tend to already be on high-speed plans that are associated with higher usage allowances. These subscribers select plans with an average speed of over 16 Mb/s and have an above-average mean of preference for speed, κ_{2h}. Thus, when UBP is imposed, these subscribers already have a large usage allowance due to their preference for speed.

The implications of UBP for consumers in category (2), for which the effect on usage and total surplus is ambiguous, are presented in Columns 3 and 4 of Table 7. In this counterfactual, all subscribers who switch plans choose more expensive plans with a larger allowance and higher speed. Overall, per-subscriber surplus decreases by \$15.83, but the ISP's revenues increase by \$21.23. Thus, the monthly per-subscriber total surplus generated by usage increases by \$5.40. Interestingly, despite the increase in surplus, usage actually decreases from 36.47 to 34.74 GBs. Columns 3 and 4 in the bottom panel of Table 7, which break down the types in category (2) who increase and decrease usage, respectively, help explain these numbers. We find that about nine out of ten subscribers in category (2) increase their usage. Those who do so have an average shock (μ_h) and an opportunity cost of content (κ_{1h}) higher than the overall averages (1.06 and 5.68, respectively), and an average preference for speed (κ_{2h}), 5.12, nearly identical to the overall average of 5.13. As a result, this group has high demand and a relatively high total disutility of content. They consume a high amount of content, they are not so sensitive to overage prices, and their waiting costs fall significantly when they choose a plan with higher speed. Hence, when they choose higher speed they consume more content.

In contrast, the average shock (0.49) and average opportunity cost of content (3.56) of the remaining one-tenth of subscribers in category (2) are less than half of the average across all types. The average preference for speed, 3.56, is also well below average. These subscribers consume little content, are more sensitive to price,[14] and their waiting costs decline comparatively less with additional speed. Hence, when they choose higher speed they consume less content. The reduction in usage from this latter group is so large that it overwhelms the increase in usage from the first group.

The effect of UBP on category (3) types is presented in Columns 5 and 6 of Table 7. Since

[14]This is reinforced by the fact that they also have a lower average utility curvature β_h, 0.38 vs. 0.41.

these types only subscribe when service is unlimited, the surplus they generate is lost when UBP is implemented. On average, this represents a loss of surplus of $44.83 per subscriber, and a reduction in usage of 23.74 GBs. From the bottom half of Table 7, it is clear that these types have a low opportunity cost of time and place little value on the content they consume.

Aggregating across the three categories of consumers, we find that total surplus generated from usage increases by $1.05 per subscriber. Additionally, per-subscriber average usage falls by 4.90 GBs. Thus, while there is significant heterogeneity in the impact of UBP across subscriber types, on average UBP modestly increases total surplus generated from usage, while transferring some surplus from subscribers to the ISP. Assuming the ISP's network costs are proportional to usage, these costs fall by more than 10%. Hence, if we could quantitatively account for network costs, our finding of improved welfare from UBP would not change.

In the above calculation we do not permit the ISP to re-optimize the prices of plans, the speeds or the number of plans.[15] It is therefore likely that we overstate the surplus lost by subscribers and the revenue gains to the ISP from UBP. One (computationally-feasible) way to address this issue and explore the sensitivity of our estimates of the welfare implications of UBP is to keep the number of plans and speeds constant, but allow the ISP to choose different fixed fees. We find that to maximize revenue associated with unlimited service, the ISP significantly raises fixed fees, by an average of $47.21. We also find that 13.3% of consumers who would subscribe under UBP plans are excluded. Furthermore, the surplus of the remaining subscribers is lower. In total, the average per-subscriber monthly surplus is higher by $53.67 under UBP and per-subscriber monthly revenues are lower by $51.65. Since traffic levels are similar, we estimate total welfare from UBP relative to this counterfactual to be unchanged. Yet the division of surplus is now much more heavily weighted towards the ISP. Thus, relative to this counterfactual, consumers strictly prefer UBP.

6.2 Economic Viability of Next-Generation Networks

We now turn to our second counterfactual, in which we evaluate welfare under the menu of UBP plans offered in the data, relative to a situation where subscribers are presented with a single plan with unlimited usage and a one Gb/s connection. This is one alternative to UBP, which can only be provided by next-generation high-speed broadband networks, that some public and private entities have embraced to deal with rising Internet traffic levels.[16] "Fiber to the home" (e.g. Google Fiber) offers essentially limitless bandwidth to subscribers, while DOCSIS 3.1 for

[15] Alternatives to service with this ISP are assumed to be unchanged. The primary alternative is DSL, which has a maximum speed of less than half the speeds offered by this ISP.

[16] Australia has embraced this solution, as has Google Fiber.

cable-broadband networks is expected to be capable of similar performance.

This counterfactual relates to the counterfactual in the previous section in several ways. First, in the above we hold the set of plans fixed, while now we evaluate one particular new plan. Second, this plan has been offered as an alternative solution for rising traffic levels. Finally, this counterfactual allows us to highlight the importance of the preference for speed in generating value.

Google Fiber currently offers one Gb/s (1,024 Mb/s) service for only $70, which is by far the cheapest of all such providers in the US.[17] Public entities with similar Gigabit service typically charge well over $100. For example, the Chattanooga Electric Power Board charges $300 for Gigabit service. Thus, Google is either drastically more efficient than other providers of such service, or it is subsidizing the service for other reasons (e.g., collecting detailed customer usage data to better understand demand).[18] For this reason, we consider a single plan with unlimited usage, a one Gb/s connection, and a fixed fee of $100. The size of the fixed fee will not alter behavior of subscribers given the unlimited nature of the service, but rather determines who subscribes and how surplus is distributed between the ISP and subscribers.

Evaluating welfare under this counterfactual scenario is simpler than the first counterfactual. Only two categories of consumers must be evaluated, those who would subscribe both to the Gigabit service and the usage-based plans (about 84.4% of the types), and those who would only subscribe to one or the other (about 9.4% of the types). For the two types of consumers, a comparison of welfare and usage is presented in the top panel of Table 8, while the bottom panel describes the types in each category.

For those types who subscribe in either case, welfare and usage for the Gigabit and UBP services are reported in Columns 1 and 2, respectively. When the unlimited Gigabit service is priced at $100, these types have surplus of $235.97 and monthly usage of 102.24 GBs. Under the menu of UBP plans, consumer surplus is $160.37 and usage is 41.80 GBs. The large increase in usage is not surprising since usage is highly elastic with respect to speed, and since Gigabit service is unlimited. The ISP's per-subscriber revenues are also lower, $70.03 versus $100. Thus, the total welfare created by these types' usage is $105.57 lower under UBP, while usage is 60.44 GBs lower. Similar to Table 7, Columns 1 and 2 in the bottom half of Table 8 describe the approximately 84% of subscriber types by whether they increase or decrease usage under UBP. Since every aspect of the Gigabit service is superior to the best UBP plan, usage is always lower

[17] One exception is Vermont Telephone Company (VTEL), which offers Gigabit service at $35 per month. However, VTEL received $94 million in federal stimulus funds to lay 1,200 miles of fiber.

[18] AT&T and others have argued that Google has been granted much more favorable terms and conditions in terms of permitting, licensing, and investment incentives.

Table 8: *Next-Generation Network Counterfactual, Usage and Welfare*

	Both		Single	
	Unlimited	**UBP**	**Unlimited**	**UBP**
	(1)	**(2)**	**(3)**	**(4)**
Percent of Types (%)	84.40		0	9.36
Usage and Surplus				
Speed (Mb/s)	1,024.00	13.97	–	11.74
Usage (GBs)	102.24	41.80	–	31.30
Consumer Surplus ($)	235.97	160.37	–	22.64
Revenue ($)	100.00	70.03	–	50.46
Total Surplus ($)	335.97	230.40	–	73.10
Δ in Total Surplus ($)	-105.57		–	73.10
	Both		**Single**	
	$\uparrow$ Usage	**$\downarrow$ Usage**	**Unlimited**	**UBP**
	(1)	**(2)**	**(3)**	**(4)**
Percent of Types (%)	0	84.40	0	9.36
Mean Type				
mean of shocks ($\bar{\mu}_h$)	–	1.26	–	-0.01
s.d. of shocks ($\bar{\sigma}_h$)	–	0.82	–	0.55
opp cost of content ($\bar{\kappa}_{1h}$)	–	6.17	–	3.82
pref for speed ($\bar{\kappa}_{2h}$)	–	5.69	–	2.86
curvature ($\bar{\beta}_h$)	–	0.40	–	0.56

Note: These statistics reflect a comparison between two counterfactual scenarios. In the unlimited scenario the consumers are offered a single plan with unlimited usage and a fast connection. In the UBP scenarios the consumers face the same set of plans, with the same characteristics, as we see in the data. Although the ISP in our data does offer unlimited plans, the characteristics of those plans are different from the hypothetical characteristics of unlimited plans considered in this counterfactual. Except for the percentages, all numbers are per-subscriber averages.

under UBP. These types tend to place above-average value on content (i.e., high μ_h) and have an above-average preference for speed (i.e., high κ_{2h}).

Intuitively, we find no types that would subscribe to Gigabit service but not the UBP menu of plans. Column 4 in the top panel of Table 8 presents the surplus and usage information for the more than 9% of subscribers served under the UBP menu of plans but not the Gigabit service. These subscribers have monthly usage of 31.30 GBs under UBP, and generate total surplus of $73.10. From the bottom panel of Table 8, it is clear that these subscribers value content and speed less.

Aggregating across the two categories of types, the total surplus generated by usage is $87.73 higher per subscriber with the Gigabit service, while monthly usage is 51.28 GBs higher. The social desirability of the investments required to provide Gigabit service then depends on whether the capital costs associated with providing the service, and any additional costs incurred from handling the traffic, can be recovered from surpluses that are $87.73 higher. However, since the ISP only realizes $21.94 in additional revenue per subscriber, there is a significant gap between social and private incentives to invest. This is consistent with over $200 million in subsidies received by Vermont's VTel and the Chattanooga's EPB from the American Recovery and Reinvestment Act of 2009 (a.k.a., Stimulus Act) to build fiber networks.

Kirjner and Parameswaran (2013) provide estimates of the capital costs associated with the fiber network being built in Kansas City by Google Fiber. The authors estimate that it will cost $84 million dollars to "pass" (i.e., run fiber along the street) 149,000 homes, or approximately $564 per household. To actually connect each home, the authors estimate it will cost Google Fiber an additional $464 per subscriber. If one assumes a 20% penetration rate for the service, this equates to capital costs of $3,284 (5*$564+$464) per household served.[19]

From a societal perspective, ignoring the costs associated with the additional traffic Gigabit service generates, these capital costs (in a metro area with population density similar to Kansas City) can be recovered in approximately 37 months ($3,284/$87.73). Ignoring additional traffic costs, and any other additional variable costs associated with operating a fiber network, makes the 37 month estimate a conservative lower bound.[20] From the ISP's perspective, the capital costs of such investment would be recovered in approximately 150 ($3,284/$21.94) months. Similarly, this estimate is a lower bound on the actual time required. If the costs associated with the additional traffic induced by Gigabit service are not prohibitively high, our estimates suggest that recovery of the costs associated with fiber networks is possible, at least in metro areas no

[19]Kirjner and Parameswaran (2013) estimate the penetration rate will be 18% by the end of the first year the service is offered.

[20]Over time, increasing demand and a preference for speed will make this a less conservative bound.

less densely populated than Kansas City. Yet the large gap between social and private incentives to invest implies that such investment will be made later than is socially optimal.

As with the first counterfactual, our estimates are based on the features of the usage-based pricing plans currently offered by the ISP providing our data. Yet it is possible to offer rather substantial improvements in network performance out of existing networks. For example, a number of cable operators are already offering speeds of 100 Mb/s on DOCSIS 3.0 networks. In this way, our estimates of gains to investment in fiber networks are an upper bound, since existing networks can be leveraged further.

The Internet also continues to play an increasingly important role in individuals' lives and traffic is growing rapidly. Our estimates of the benefit to subscribers is based on current demand, and can change dramatically with the development of a single application. Thus, our estimates will need to be refreshed fairly often to be of value. Finally, the fixed fee we use is somewhat arbitrary and may reflect substantial subsidies. When we allow the ISP to maximize the revenues associated with providing unlimited Gigabit service, the optimal fixed fee is $194.65 and 83.91% of subscriber types are served. Thus, only a few marginal types are no longer served and the higher fixed fee simply represents a transfer from consumers to the ISP, leaving total welfare largely unchanged. This revenue-maximizing price is in the middle of the range of prices currently offered for Gigabit service in the US, giving us further confidence in our demand estimates. With this higher fixed fee the capital costs would be recovered in just over 31 months. However, due to restrictions on rates from local municipalities, an ISP may have a difficult time charging this rate.

7 Conclusion

We use high-frequency usage data, and variation in the shadow price of usage, due to usage-based pricing, to estimate demand for residential broadband service. We then use the estimates to evaluate the welfare implications of two alternatives proposed to dealing with network congestion: usage-based pricing and high-speed next-generation networks.

Our results suggest that usage-based pricing is an effective means to remove low-value traffic from the Internet, while improving overall welfare. Consumers adopt higher speeds, on average, which lowers waiting costs. Yet overall usage falls slightly. The effect on subscriber welfare depends on the alternative considered. If we hold the set of plans, and their prices, constant, then usage-based pricing is a transfer of surplus from consumers to ISPs. However, if we let the ISP set price to maximize revenues, then consumers are better off.

An alternative to usage-based pricing which can provide near-limitless bandwidth to subscribers has made the industry's move toward usage-based pricing controversial. Our estimates allow us to evaluate the incentives for ISPs to invest in high-speed next-generation networks capable of delivering speeds of one Gb/s. The significant heterogeneity in demand for plan features results in nearly half of subscribers actually preferring a menu of usage-based pricing plans to a single high-speed, and by our estimates heavily subsidized, plan similar to that offered by Google Fiber. However, this is offset by more than 20% of subscribers whose surplus would increase substantially ($50 or more) from such a plan, and higher revenues for the ISP. This makes cost recovery for the ISP possible; however, the ISP's inability to capture the additional surplus to subscribers yields a large gap between the private and social incentives for investment in such networks. This suggests that without subsidization these investments will come much later than is socially optimal.

Network congestion, which is argued to be a driver of the move towards usage-based pricing, was not necessary to model because the ISP providing our data operated an overly-provisioned network. An interesting question for future research is to measure the size and impact of congestion externalities among subscribers. This question is better studied on a network that suffers from congestion, especially if this congestion varies with network upgrades.

REFERENCES

Abaluck, Jason and Jonathan Gruber. 2012. "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program", *Forthcoming: American Economic Review.*

Ackerberg, Daniel. 2009. "A New User of Importance Sampling to Reduce Computational Burden in Simulation Estimation", *Quantitative Marketing and Economics*, 7(4), 343-376.

Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen. 2012. "Moral Hazard in Health Insurance: How Important Is Forward-Looking Behavior?", *NBER Working Paper No. 17802.*

Bajari, Patrick, Jeremy Fox, and Stephen Ryan. 2007. "Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients", *American Economic Review P&P*, 97(2), 459-463.

Bauer, Johannes M., and Steven Wildman. 2012. "The Economics of Usage-Based Pricing in Local Broadband Markets", *Michigan State University Working Paper.*

Chevalier, Judith, and Austan Goolsbee. 2009. "Are Durable Goods Consumers Forward-Looking? Evidence from College Textbooks", *Quarterly Journal of Economics*, 124(4), 1853-1884.

Chung, Doug, Thomas Steenburgh, and K. Sudhir. 2010. "Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus-Based Compensation Plans", *HBS Working Paper 11-041.*

Copeland, Adam, and Cyril Monnet. 2009. "The Welfare Effects of Incentive Schemes", *Review of Economic Studies*, 76(1), 93-113.

Crawford, Gregory, and Matthew Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand", *Econometrica*, 73(4), 1137-1173.

Dutz, M., J. Orszag, and R. Willig. 2009. "The Substantial Consumer Benefits of Broadband Connectivity for US Households", Mimeo.

Edell, Richard and Pravin Varaiya. 2002. "Providing Internet Access: What We Learn from INDEX", Washington, DC, Editors: James Alleman and Robert Crandall, Volume: "Broadband: Should We Regulate High-Speed Internet Access?", *Brookings Institution.*

Fox, Jeremy, Kyoo il Kim, Stephen Ryan, and Patrick Bajari. 2011. "A Simple Estimator for the Distribution of Random Coefficients", *Quantititative Economics*, 2(3), 381-418.

Goettler, Ronald, and Karen Clay. 2011. "Tariff Choice with Consumer Learning and Switching Costs", *Journal of Marketing Research*, 48(4), 633-652.

Goolsbee, Austan, and Peter J. 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.

Grubb, Michael and Matthew Osborne. 2012. "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock", *MIT Working Paper.*

Grubb, Michael. 2012. "Consumer Inattention and Bill-Shock Regulation", *MIT Working Paper.*

Gowrisankaran, Gautam, and Marc Rysman. 2012. "Dynamics of Consumer Demand for Durable Goods", *Journal of Political Economy*, 120(6), 1173-1219.

Handel, Benjamin. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts", *Forthcoming: American Economic Review.*

Hendel, Igal, and Aviv Nevo. 2006a. "Measuring the Implications of Sales and Consumer Inventory Behavior", *Econometrica*, 74(6), 1637-1673.

Hendel, Igal, and Aviv Nevo. 2006b. "Sales and Consumer Inventory", *The RAND Journal of Economics*, 37(3), 543-561.

Hitt, L. and P. Tambe. 2007. "Broadband Adoption and Content Consumption", *Information Economics and Policy*, 19(3-4), 362-378.

Kirjner, Carlos and Ram Parameswaran. 2013. "Google Fiber - How Much Will Google Spend to Overbuild Kansas City? How Much Would It Cost to Expand?", *Bernstein Research.*

Lahiri, S.N. 2003. "Resampling Methods for Dependent Data", *Springer.*

Lambrecht, Anja, Katja Seim, and Bernd Skiera. 2007. "Does Uncertainty Matter? Consumer Behavior Under Three-Part Tariffs", *Marketing Science*, 26(5), 698-710.

Mackie-Mason, Jeffrey and Hal Varian. 1995. "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas in Communications* 13, 1141-49.

Misra, Sanjog and Harikesh Nair. 2011. "A Structural Model of Sales-Force Compensation Dynamics: Estimation and Field Implementation", *Quantitative Marketing and Economics*, 9(3), 211-257.

Nielsen. 2012 *The Cross-Platform Report,* Quarter 2, 2012. Downloaded from http://www.nielsen.com/us/en/insights/reports-downloads/2012/state-of-the-media–

cross-platform-report-q2-2012.html.

Odlyzko, Andrew, Bill St. Arnaud, Erik Stallman, and Michael Weinberg. 2012. "Know Your Limits: Considering the Role of Data Caps and Usage Based Billing in Internet Access Service", Whitepaper, *Public Knowledge.*

Open Internet Advisory Committee (OIAC). 2013. "Policy Issues in Data Caps and Usage Based Pricing", OIAC 2013 Annual Report, 14-38. Available at: http://transition.fcc.gov/cgb/oiac/oiac-2013-annual-report.pdf.

Rainie, Lee and Dan Packel. 2001. "More Online, Doing More," Pew Internet and American Life Project 3, available at http://www.pewinternet.org/Reports/2001/More-Online-Doing-More.aspx.

Rappoport, Paul, Donald Kridel, and Lester Taylor. 2002. "The Demand for Broadband: Access, Content, and the Value of Time", Washington, DC, Editors: James Alleman and Robert Crandall, Volume: "Broadband: Should We Regulate High-Speed Internet Access?", *Brookings Institution.*

Rosston, G., S. Savage, and D. Waldman. 2010. "Household Demand for Broadband Internet in 2010", *The B.E. Journal of Economic and Policy Analysis (Advances)*, 10(1), Article 79.

Thaler, Richard and Cass Sunstein. 2008. "Nudge: Improving Decisions About Health, Wealth, and Happiness", New York, NY, *Penguin Books.*

Varian, Hal. 2002. "The Demand for Bandwidth: Evidence from the INDEX Experiment", Washington, DC, Editors: James Alleman and Robert Crandall, Volume: "Broadband: Should We Regulate High-Speed Internet Access?", *Brookings Institution.*

Yao, Song, Carl Mela, Jeongwen Chiang, and Yuxin Chen. 2012. "Determining Consumers' Discount Rates with Field Studies", *Journal of Marketing Research*, 49(6), 822-841.