NBER WORKING PAPER SERIES

USAGE-BASED PRICING AND DEMAND FOR RESIDENTIAL BROADBAND

Aviv Nevo
John L. Turner
Jonathan W. Williams

Working Paper 21321
http://www.nber.org/papers/w21321

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2015

We are grateful to the North American Internet Service Provider that provided the data used in this paper. We thank participants in several seminars, Gautam Gowrisankaran, Shane Greenstein, Tom Holmes, Nathan Miller, Mar Reguant, 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). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Usage-Based Pricing and Demand for Residential Broadband
Aviv Nevo, John L. Turner, and Jonathan W. Williams
NBER Working Paper No. 21321
July 2015
JEL No. L1,L13,L96

ABSTRACT

We estimate demand for residential broadband using high-frequency data from subscribers facing a three-part tariff. The three-part tariff makes data usage during the billing cycle a dynamic problem; thus, generating variation in the (shadow) price of usage. 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 the estimates, we simulate demand under alternative pricing and find that usage-based pricing eliminates low-value traffic. Furthermore, we show that the costs associated with investment in fiber-optic networks are likely recoverable in some markets, but that there is a large gap between social and private incentives to invest.

Aviv Nevo
Department of Economics
Northwestern University
2001 Sheridan Road
Evanston, IL 60208-2600
and NBER
nevo@northwestern.edu

John L. Turner
Department of Economics
University of Georgia
jlturner@terry.uga.edu

Jonathan W. Williams
Department of Economics
107 Gardner Hall, CB 3305
University of North Carolina
Chapel Hill, NC 27599-3305
jonwms@unc.edu

1 Introduction

The telecommunications sector is undergoing major changes and is the focus of several important public policy debates. A driving force behind these changes is the growing importance of data services and the proliferation of online activities, especially the popularity of (over the top) video providers such as Netflix and YouTube. Cable companies, which once mainly delivered video, are shifting their focus to broadband services. The same is true for cellular carriers, whose networks are increasingly used to deliver data. Traditional telecomm companies are trying to keep up with this trend and offering their version of data delivery services.[1] In this paper, we contribute an important ingredient for studying the economics of this industry: demand for residential broadband. In particular, we estimate demand using a unique dataset and (shadow) price variation created by usage-based pricing. We demonstrate the implications of the demand estimates by computing consumers' plan choices and usage when faced with a variety of contracts including unlimited plans and high speed fiber-to-the-premise (FTTP) options.

In order to estimate demand for residential broadband we rely on two sources of variation. We use variation in prices and attributes of the plans subscribers choose, but more importantly we rely on variation created by three-part tariff plans we observe. Subscribers pay a monthly fee, which provides them a monthly data allowance. If they exceed the allowance they pay a price per Gigabyte (GB). When facing a three-part tariff plan, the marginal price paid for usage is zero until the subscriber exceeds their allowance. However, a forward-looking subscriber realizes that the shadow price of usage 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.

At the core of the paper is a dataset that we secured from an Internet Service Provider (ISP). The data include information on hour-by-hour Internet usage for roughly 55,000 subscribers facing different price schedules. We also know plan-specific variables (speed, prices, etc.) for the plan the household is subscribed to and for the alternatives not chosen. The ISP has in place three-part tariff plans in addition to subscribers who are grandfathered in to unlimited plans.

Using these data we provide descriptive evidence that consumers respond to variation in the

[1] Major public policy discussions in this sector include the proposed merger between two of the largest cable and broadband providers in the US, Comcast and Time Warner Cable, the proposed merger between ATT and DirecTV, which claim, in public filing with the FCC, that the merger will generate synergies between ATT's broadband abilities and DirecTV's video. Both these mergers are currently being evaluated by the DOJ and FCC. A third proposed merger, between Sprint and T-Mobile, was canceled but was rumored to be driven by incentives to invest in next generation wireless technology that will deliver faster data delivery services. Other policy issues facing the FCC and Congress are the equal treatment of content on the Internet (so called net neutrality) and encouragement of municipal broadband networks.

shadow price of usage. We then estimate a (finite horizon) dynamic choice model, by adapting the techniques of Ackerberg (2009), Bajari et al. (2007) and Fox et al. (2011). Specifically, we solve the dynamic problem for a large number of subscriber types, once for each type. We then estimate the distribution over these types by matching moments recovered from the data to those predicted by a weighted average of the optimal behavior of the types.

The estimates allow us to quantify several aspects of consumer demand for broadband. First, we find that consumers' willingness-to-pay for a one Megabit per second (Mb/s) increase in connection speed is between \$0 to \$5 per month, with an average of \$2.03. Second, we find that the average consumer's willingness-to-pay for a one GB increase in usage allowance is \$0.36 per month, which suggests that marginal content has relatively low value. However, at a price of zero and with average download speed, we find the average subscriber's usage would average about 66 GB per month and consumer surplus would average nearly \$165 per month.

To demonstrate the implications of the demand estimates, we calculate usage and welfare under several alternative menus of plans. We find that usage-based pricing is effective in lowering usage, without reducing consumer welfare significantly, because it mostly removes content with relatively low value. Generally, usage-based pricing shifts surplus from consumers to providers. The magnitude, as well as the effect on total welfare, depends on how the prices of the unlimited plans are set.[2]

Next we evaluate adoption, usage and welfare when consumers are presented with an unlimited plan and a one Gigabit per second (Gb/s) connection. We find that surplus generated from usage is substantial. Yet at a fee of \$70, which is what Google charges for Google Fiber in Kansas City, the ISP captures only a small portion of this surplus, generating a gap between private and social benefits from investment. Using cost estimates from Kirjner and Parameswaran (2013), we estimate it takes 27 months to recover the capital expenditures from a social perspective, relative to typical cable offerings, and only 12 months relative to not having any broadband service. On the other hand, a typical ISP that upgrades to Gigabit speeds would recover these costs only after about 149 months. The exact recovery time depends on the competitive environment, but

[2] Usage-based pricing has been proposed as one way to manage congestion in the current bandwidth-intensive environment. The term typically refers to non-linear pricing based on the quantity of usage and not the type, or content, of usage, which is at the heart of the net-neutrality debate. Usage-based pricing is popular for broadband 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 service (OIAC 2013). Much of the debate on the usage and 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. See De Fontenay et al. (1990) for a discussion of similar issues with telecommunications services in the 1980s. Another way to manage congestion is setting peak-load prices and giving users incentives to shift usage to off peak times. Without a change in some of the major applications that currently require real-time streaming, this strategy is unlikely to be effective.

the general point – that there is a large gap between social and private incentives to invest – is quite robust.

Our paper is related to a literature that studies demand for broadband service. Varian (2002) and Edell and Varaiya (2002) run experiments, where users face varying prices for different allowances and speeds. Goolsbee and Klenow (2006) use data on individuals' time spent on the Internet and earnings to estimate consumer benefit from residential broadband, assuming an hour spent on the Internet is an hour of forgone wages. 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 using plan choice data. 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 several literatures. First, is a literature that focuses on estimating demand in dynamic settings (Crawford and Shum, 2005; Hartmann, 2006, Hendel and Nevo, 2006a; Gowrisankaran and Rysman, 2012; and others). Like our analysis, Yao et al. (2012) exploits intra-month (weekly) variation in the shadow price of usage under three-part tariffs to identify consumers' discount factors. Second, is a literature studying incentives in labor contracts where a nonlinear compensation structure based on performance during a fixed period of time 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 (e.g. Copeland and Monnet (2009), Chung et al. (2010), and Misra and Nair (2011), Einav at al (forthcoming)). Finally, our paper is related to a literature that examines if consumers are forward-looking (Aron-Dine et al. (2012), Chevalier and Goolsbee (2009), Grubb (2015), Grubb and Osborne (2015), and Hendel and Nevo (2006b)).

2 Data

The data come from a North American ISP that offers several plans. Features of a plan include maximum download speed, an access fee, usage allowance (if any), and overage price per GB of data (if any).[3] Unlimited plans, where subscribers do not face overage prices, are only available to subscribers who previously had them. Usage in GBs is recorded for both uploads and downloads, but for billing purposes, and consequently our purposes, the direction of the traffic is ignored. For each subscriber, we observe usage at the monthly level from May 1^{st}, 2011 to May 31^{st}, 2012, and for 15-minute intervals during May 10^{th} to June 30^{th}, 2012. We also know the plan chosen

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

by the subscriber.[4]

2.1 Sample and Descriptive Statistics

The sample includes 54,801 subscribers in four different markets served by the 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.

The data demonstrate a sharp increase in usage. The median subscriber more than doubles usage, from 9 GBs in May 2011 to over 20 GBs in May 2012, an increase equivalent to four high-definition movies. The average subscriber's usage increases from 23 to over 40 GBs. There is substantial heterogeneity in usage: the 25th percentile household used less than 6 GB per month in May 2012, while the 75th percentile consumed almost 9 times as much. During this period, approximately 39% of aggregate traffic is online video, 35% is web traffic, 14% is peer-to-peer activities, and the remainder is largely comprised of gaming activities, software updates, and cloud-based and music-streaming services. These and other patterns are presented in the Appendix.

Table 1: *Descriptive Statistics of Subscriber Plan Choices and Usage, May-June 2012*

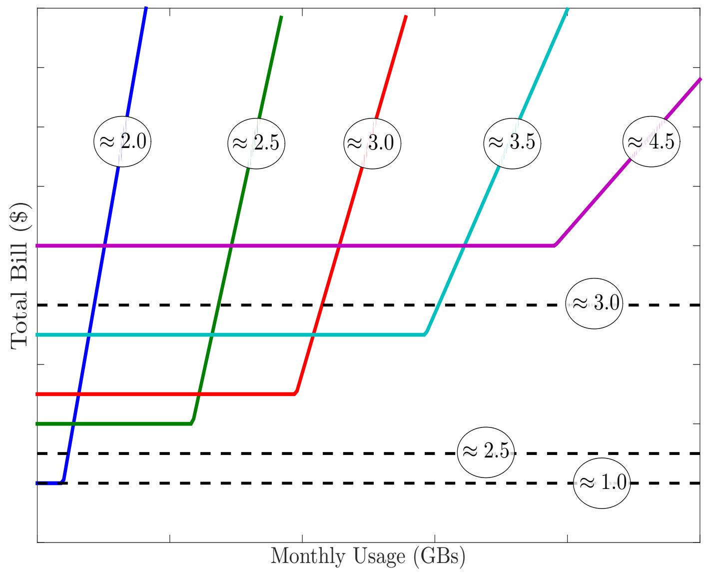
	Unlimited Plans	**Usage-Based Plans**
Number of Subscribers	12,316	42,485
Plan Characteristics		
Mean Access Fee ($)	44.33	74.20
Mean Download Speed (Mb/s)	6.40	14.68
Mean Allowance (GB)	∞	92.84
Mean Overage Price ($/GB)	0.0	3.28
Usage		
Mean (GB)	50.39	43.39
Median (GB)	25.60	23.63
Median Price per 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 May-June 2012. 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.

For the main analysis we use disaggregated data, which include one complete billing cycle for each subscriber during May and June, 2012. Usage shows a cyclical pattern during the

[4] 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.

Figure 1: *Plan Features and Billing*



Note: This figure illustrates the relationship between monthly usage and cost for the usage-based and grandfathered plans. The approximate relative speed for each plan, normalized by the slowest plan, is indicated by the circle intersecting each line.

day. Peak usage occurs 10pm-11pm, when the average user consumes over 4.5 GBs each month. This is almost six times the amount of traffic generated during 5am-6am. Throughout the day, approximately 90% of the usage is for download. Aggregating usage (uploads and downloads) to a daily level results in 1,644,030 subscriber-day observations.

Table 1 reports summary statistics on monthly usage and plan characteristics for May-June 2012 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-based plan pays nearly $30 more per month to enjoy faster download speed (14.68 Mb/s), but uses under 44 GB. The median subscriber to an unlimited plan pays about $1.68 per GB, less than 60% of what a usage-based subscriber pays.

Our data use agreement prevents us from disclosing the actual plan features, but we can show an approximate relative ranking of the plan features and costs. For each plan, Figure 1 presents how the total cost of subscribing to a plan changes with monthly usage. Each plan is also labeled with a relative speed ranking, with the slowest plan normalized to 1. The presence of low-cost, low-speed, and unlimited-usage grandfathered plans allows us to better characterize the shape of the utility function by observing a marginal price for usage equal to zero.

In Table 2, we report summary statistics for overage charges incurred by subscribers on

Table 2: *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 that would have yielded a lower cost and with download speed that is no lower. Means and medians are at the subscriber level.

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 with a positive 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 GBs. In the Appendix we provide the distribution of the fraction of allowance used.

Previous work has studied whether consumers make what look like suboptimal choices ex-post in a variety of settings.[5] In our data, it is not obvious that subscribers systemically make mistakes. One way to measure these mistakes is to ask how many subscribers could switch to a plan that costs less and is no slower. Using this definition, if we look at the complete billing cycle (June 2012) in isolation, we find 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. This is a weak test of the optimality of plan choice since, as we see in Figure 1, speed and allowance are positively correlated. So, for example, if we see a consumer use little of their allowance we cannot be certain whether they have a high willingness to pay for additional speed or if they chose a sub-optimal plan.

[5]In particular, some research has highlighted what seem like suboptimal choice made by by consumers facing non-linear pricing, similar to ours, in cell phone usage (Grubb and Osborne 2015) and health care (Abaluck and Gruber 2011; Handel 2013). On the other hand, several papers (e.g., Miravete 2003; Economides et al 2008; Goettler and Clay 2011; Kecham et al 2012) highlight circumstances where individuals make choices that are rational ex post.

2.2 Are Subscribers Forward Looking?

We now examine whether consumers in our data are forward looking. This is interesting for two reasons. First, the evidence we provide adds to a growing literature demonstrating consumers are forward-looking when making economic choices. 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 model.

Our data are from an ISP 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 cumulative usage during the month.

If subscribers are forward looking, we expect certain patterns in usage throughout a billing cycle. The heaviest-volume subscribers who know they have high probability of exceeding their allowance should behave as though the shadow price is equal to the overage price: there should be little change in average usage throughout the billing cycle. Similarly, for subscribers with only a small probability of exceeding their allowance, 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

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

where $ln(c_{jkt})$ 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 that have used a substantial proportion of their allowance by this time. We include include dummy

Table 3: *Forward-Looking Behavior, Within-Month Regression*

	$\mathbb{1}\left[10 \leq t < 15\right]$	$\mathbb{1}\left[15 \leq t < 20\right]$	$\mathbb{1}\left[20 \leq t < 25\right]$	$\mathbb{1}\left[25 \leq t < 31\right]$
$1\left[0 \leq \frac{C_{jk(t-1)}}{\overline{C}_k} < 0.40\right]$	-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			

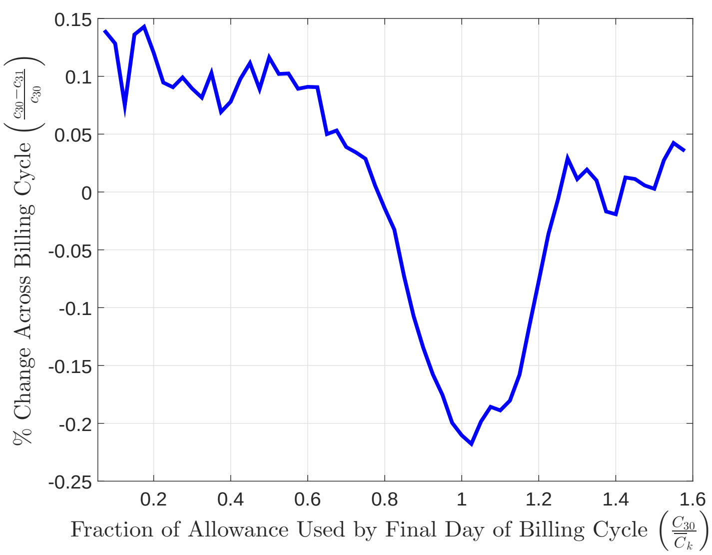
Note: This table presents OLS estimates of Equation (1) using 1,644,030 subscriber-day observations. The dependent variable is natural logarithm of daily usage. Each cell in the table gives the coefficient on the interaction between the indicators in the respective row and column. Controls include a constant, time trend, indicators for the day of the week, and subscriber fixed effects. Asterisks denote statistical significance: ** 1% level, * 5% level.

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. Since different households have different billing dates we can separate time trends and demand shocks on certain days from the dynamics of our model. We also include subscriber fixed effects, μ_j, to remove persistent forms of heterogeneity across subscribers.

The estimates of Equation (1) are reported in Table 3. Each cell reports the estimate for the coefficient on the interaction between the indicators in the respective row and column. The patterns in the table are consistent with forward-looking behavior: at each point in the billing cycle current usage is lower the closer the consumer is to the allowance (and hence the shadow price is higher). Furthermore, subscribers who are near the allowance early in the billing cycle reduce usage less than subscribers near the allowance later in the billing cycle (i.e., coefficients decrease monotonically from left to right within rows four and five of Table 3). This is consistent with consumers near the allowance later in the billing cycle reducing usage proportionally more, relative to their own mean, than consumers 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.

In addition to the within-month variation in price, subscribers also encounter a change in the shadow price when their usage allowance is refreshed at the beginning of a new billing cycle. A forward-looking subscriber near the allowance at the end of a billing cycle knows that the shadow price 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 2: *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, $\frac{c_{jk(30)}-c_{jk(31)}}{c_{jk(30)}}$. 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 2 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.[6]

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

[6]In the Appendix we provide further analysis to demonstrate that the result holds for different time windows.

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 From Content

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 content is delivered, s_k, by a usage allowance, $\overline{C}_k$, by a fixed fee F_k that pays for all consumption up to the usage allowance and by an overage price, p_k per GB of usage in excess of the allowance. 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.[7] 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. The utility for a subscriber of type h on plan k is given by

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

The first term captures the subscriber's gross utility from usage. Quickly declining marginal utility seems natural: a subscriber's first e-mail sent, her favorite website to check, her first Netix/Youtube video, etc., should bring significantly more marginal utility than subsequent usage. The specification – where the *curvature* of the utility function can vary from log ($\beta_h \to 1$) to linear ($\beta_h = 0$) – permits an interpretation based on a close link to price elasticity of demand. The utility from consumption is scaled by a time-varying shock, υ_t, which captures randomness in utility from consumption. The shock is observed by the subscriber only in 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 0.5% 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. A more general model would relax the iid assumptions in a variety of ways. For example, the distribution could vary by day of the week, exhibit serial correlation or be more "lumpy" (to account for different usage generated, say, by email and movies). However, as we show in the Appendix, it seems like none of these are a major concern in our data.

[7] In this way, we assume content with a similar marginal utility is generated each day or constantly refreshed. This may not be the case for a subscriber who has not previously had access to the internet.

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*. 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 for explaining why consumers on unlimited plans consume a finite amount of content.[8] Our specification assumes that the cost of consuming content is linear, but the marginal utility is convex. Alternatively, we could allow for curvature in cost. However, it is not clear how the data would separate these two and it seems to us that a linear cost, reflecting time cost, and declining marginal utility is the natural assumption.

The vector of parameters, $(\beta_h, \kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$, describes a subscriber of type h. Conditional 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 model daily decisions, over a finite and short horizon. Uncertainty involves the realizations of υ_t. We assume that wealth, I, is large enough so that it does not constrain consumption of content.

3.2 Optimal Usage

We now solve for the optimal usage implied by the model. 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, 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 the preference shock, υ_T. 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 $\frac{\partial u_h(c_t, y_t, \upsilon_t; k)}{\partial c_t} = p_k$. Denoting this optimal level of consumption by $c^*_{hkT}(C_{T-1}, \upsilon_T)$,

[8] The data come from a period when the network of this ISP was overly-provisioned. Therefore, we assume that the usage decision does not depend on concerns over congestion in the network.

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. We therefore solve for the optimal usage recursively. Specifically, type h on plan k solves

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

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

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

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 $-\frac{1}{\beta_h}$. 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

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

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 implies a distribution for the time spent in particular states (t,C_{t-1}) over a billing cycle.

3.3 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. The subscriber may also choose no plan at all, $k = 0$. Formally, the plan choice is given by

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

where the value $E[V_{h01}(0)]$ and the fixed access fee F_0 for $k = 0$, the outside option, are normalized to 0. Note, that we assume that there is no error, so consumers choose the plan that is optimal. This assumption is motivated by the finding above that consumers seem to make few choices that ex-post are clearly mistakes.[9]

4 Estimation and Identification

We estimate the parameters of the model 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. Second, we estimate a weight for each of the types by matching the weighted average of optimal behavior, calculated in the first stage, to the equivalent moments observed in the data. This yields an estimated distribution of types. In this section we outline the main steps and provide more details and sensitivity analysis in the Appendix.

In step 1 of the estimation we solve the dynamic program for 16,807 types (seven points of support for each of the five parameters), where each type is defined by a value of the parameter vector $(\beta_h, \kappa_{1h}, \kappa_{2h}, \mu_h, \sigma_h)$. 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. To do so, we discretize the state space. Because the subscriber does not know υ_t prior to period t, we can integrate over its support and the solution to the dynamic programming problem for each type of subscriber can be characterized by the expected value functions, $E[V_{hkt}(C_{t-1})]$, and policy functions, $E[c^*_{hkt}(C_{t-1})]$. Having solved the dynamic program for a subscriber of type h, we generate the transition process for the state vector implied by the solution.

In step 2 of the estimation we choose a weight for each subscriber type to match moments we recover from the data to the (weighted) average of the behavior predicted by the model.

[9]In the Appendix we show how we can generalize the model along several dimensions.

Formally, we choose weights to satisfy

$$\widehat{\theta} = \arg\min_{\theta} \mathbf{m}_k(\theta)'\widehat{\mathbf{V}}^{-1}\mathbf{m}_k(\theta), \quad \text{subject to} \quad \sum_{h=1}^{H_k} \theta_h = 1 \text{ and } \theta_h \geq 0 \ \ \forall h.$$

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, $\mathbf{m}_k^{mod}\theta$ is a weighted average of the equivalent type-specific moments predicted by the model, and $\widehat{\mathbf{V}}^{-1}$ is a weighting matrix. Note that type weights, θ_h, are chosen to match the moments for each plan, and they sum up to 1 for each plan. After the estimation we rescale the weights by the probability that each plan is chosen, and therefore we also match the share of each plan in the data.

In choosing which moments to match we focus on two considerations: identification and computational ease. Estimation is much simpler if the moments are linear in the weights.[10] We therefore 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 given 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, which keeps the moment linear in the parameters. The second set of moments is the mass of subscribers at a particular state, $\sum_{h=1}^{H} \gamma_{hkt}(C_{t-1})\theta_h$, which like the first set of moments, is easy to calculate and linear in the weights.

We calculate standard errors using a block-resampling methodology (Lahiri 2003). Specifically, we sample the data by consumer with replacement, keeping all 30 days for each of 54,801 consumers drawn, which results in 1,000 samples of size 1,644,030. For each sample, we re-calculate the moments and then re-estimate the weights. We calculate standard errors for subsequent statistics and counterfactual analyses by repeating the calculation using the 1,000 different estimates of the weights.

The estimation procedure recovers the weights of each type by choosing the mixture of types that best matches the data. The data we use to identify the parameters includes plan choice and usage. The logic of identification follows in some ways that of Bayesian estimation. The selection of the boundaries of the initial grid amounts to putting a uniform prior on the distribution of types over the grid (and zero probability elsewhere). Plan selection – specifically, the share of consumers who choose each plan – provides information on the type distribution. In the model, each type has an optimal plan, so plan selection splits the type space into distinct groups but

[10] 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 as long as the moments are linear in the weights.

does not provide any information about the relative importance of types within a group. In other words, it puts a weight on each group of types equal to the share of the plan that group chooses. As the number of plans increases, and the attributes and prices of the plans vary, the type distribution can be recovered.

The usage moments allow us to distribute the weight among types within a group who choose each plan. 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: the weights are identified as long as the behavior predicted by different types is not collinear over all the moments and all states used in estimation. Thus, the key to identification is to understand how each parameter impacts the variation in predicted behavior across moments and states.

We rely on two types of usage variation. The first is variation in mean usage across different states (i.e., day in the billing cycle and fraction of allowance used). The second is variation in higher-order moments of the usage distribution. Rather than calculating a variety of such moments, we focus on the cumulative distribution of usage, which efficiently summarizes the information in the higher-order moments. Note that the first set of moments we use are the unconditional mean usage, i.e., taking into account usage conditional on getting to a state and the probability of reaching that state.

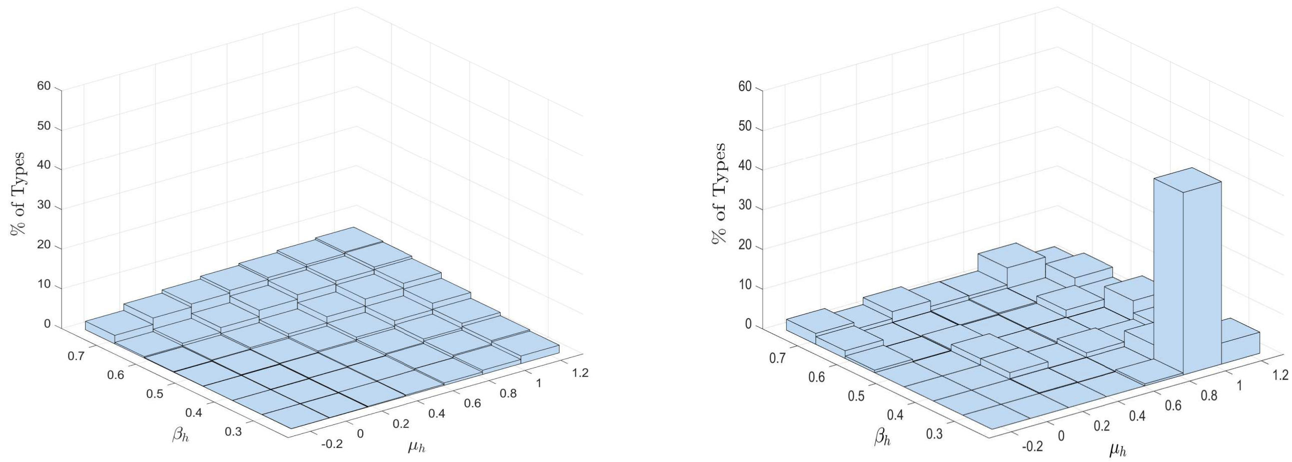
Each parameter impacts behavior in these states in a different way. For example, roughly speaking, the parameter μ_h determines average usage across days for each type. To see this, consider Equation (2), but fix $\beta_h = 1$ and assume that the shadow price does not vary with the parameters. In this case average consumption is given by

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

An increase in the average of the shocks impacts usage in all states. On the other hand, changes in the disutility of consuming content, κ_{1h} and κ_{2h}, have a different effect for different shadow prices. With $\beta_h \neq 1$ and with the shadow price varying with the parameters, the same idea holds: the change in mean usage across states is affected in a different way by the different parameters.

In addition to variation in average usage across states, we also rely on changes to higher-order moments of usage. A change in σ_h impacts the variance of usage and therefore the likelihood of reaching certain states. Other parameters also impact the probability of reaching certain states. For example, while different combinations of the curvature parameter β_h and the disutility parameters, κ_{1h} and κ_{2h}, may imply similar average usage, these parameters affect higher-order moments differently. In the absence of overage prices, a high-curvature, low-disutility subscriber could have similar average usage as a low-curvature, high-disutility subscriber. But the latter

Figure 3: *Sources of Identification: Plan Selection and Usage*



(a) *Only Plan Selection*

(b) *Plan Selection and Usage*

Note: Figure 3(a) presents the joint distribution of the utility curvature parameter, β_h, and the mean of shocks, μ_h, when only information on optimal plan selection is used and uniform weights are applied. Figure 3(b) presents the distribution when information on optimal plan selection is used and the weights are chosen to match usage moments from the data.

subscriber is relatively less sensitive to movements in the shadow price than the former. Hence, mean usage across states and the probability of reaching states will be different.

The nonlinearity of the parameters makes it hard to precisely define what variation in the data identifies each parameter. However, loosely speaking, plan selection is critical for pinning down κ_{2h}. The mean usage across all states pins down μ_h and the variance in usage across all states pins down σ_h. Finally, β_h and κ_{1h} are identified from the variation in usage across states with different shadow prices. In the Appendix we further demonstrate how types behave differently by looking at the behavior of the types we estimate.

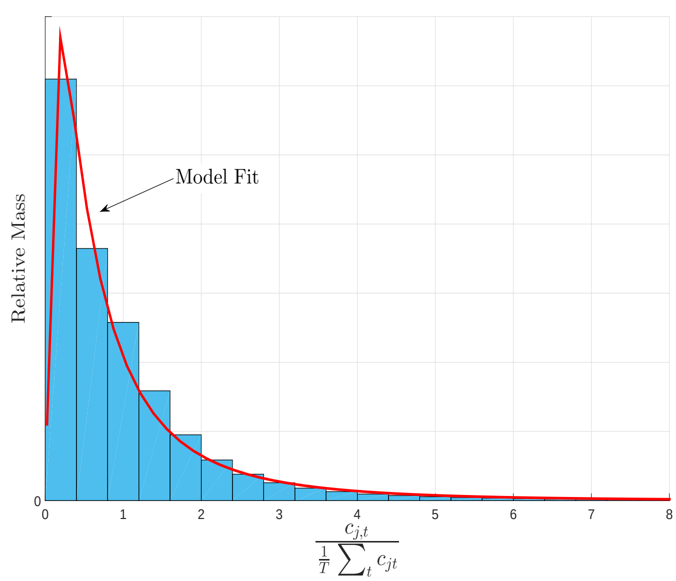
To demonstrate the importance of plan selection and the usage moments in pinning down the distribution of types, we present below the estimated distribution when we only use plan choice and when we use both plan choice and usage.

5 Results

We estimate a weight greater than 0.01% ($\theta_h > 0.0001$) for 53 types. The most common type accounts for 28% of the total mass, the top 5 types account for 65%, the top 10 for 78% and the top 20 for 90%.[11] No plan has more than 20 types receiving positive weights, while the average number of types across plans is only 6.63.[12] For the most common type, with average speed on 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 \$8.10/GB, and average daily consumption (in absence of overages) would be about 29 GB and gross willingness-to-pay would be about \$72.74.

Figure 3 demonstrates the importance of plan selection and information on usage (during the billing cycle) for identifying the distribution of types. As we noted above, in the model, each type has an optimal plan, so plan selection splits the type space into distinct groups but does not provide any information about the relative importance of types within a group. Figure 3(a) presents the joint distribution of the utility curvature, β_h, and the mean of the distribution of random shocks, μ_h, if only plan selection information is used. That is, if we put equal weight on each of the types within a group (so that the weights sum up to the share of the plan), and then integrate over the other parameters to recover the joint distribution of β_h and μ_h. Figure 3(b) presents our estimates of the joint distribution when we use both plan choice and usage information, as we discuss above. The two distributions look very different, suggesting that the usage moments are crucial source of information.[13]

Figure 4: *Model Fit: Distribution of Usage Relative to a Subscriber's Mean*



Note: This figure presents the ratio of daily usage, c_{jt}, to a subscriber's monthly average, $\frac{1}{T}\sum_t c_{jt}$, from the data and simulations from the model.

[11]The top 20 types, as well as some statistics from the estimated distribution, are presented in the Appendix.

[12]Expanding the grid of types to allow for two additional values of each parameter results in estimates that assign no weight greater than 0.01% to any of the additional types.

[13]In the Appendix, we show the marginal distribution of all 5 parameters when we use different moments for estimation.

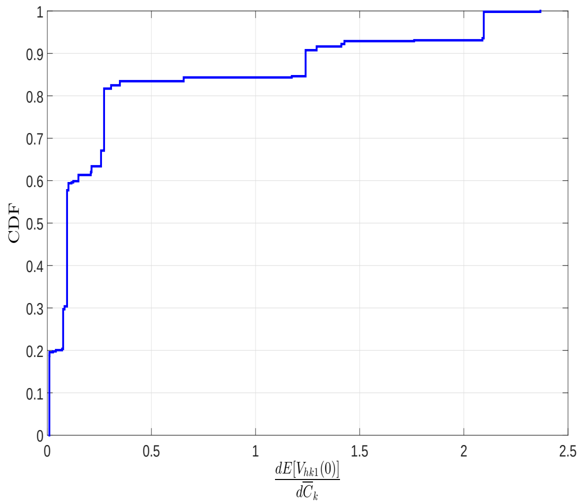
The estimated joint distribution in Figure 3(b) is highly irregular and looks quite different from a normal distribution. For the highest-volume subscribers (high μ_h), there is substantial variation in the elasticity of demand. In fact, for high-μ_h subscribers, the distribution of β_h is clearly multi-peaked (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. The distribution of the other parameters is similarly iregular and multi-peaked. For example, 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.

Overall, our model fits the data quite well. For all plans, the correlation between the empirical moments and the fitted moments is very high. The model also does well fitting patterns in the data that were not explicitly matched during estimation. For example, one might be concerned that bandwidth-intensive activities, like online video and cloud-based services, will generate a lumpiness in usage that our log-normal distribution cannot capture. To address this, we calculate in the data, and simulate from the model, the ratio of daily consumption to a consumer's mean usage over a month. In the model, this ratio will vary both because of variation in the state and because of random draws to the consumption shocks, υ_t. We therefore simulate usage predicted by the model 1,000 times for each consumer type over a 30 day period, perform the same calculation for each of the simulations, and then aggregate across types accounting for the relative mass of each. Figure 4 presents the distribution of this ratio for the data and the model simulations. Overall the model matches the data quite well, which both provides a goodness of fit measure and shows that our model can deal with whatever lumpiness in usage is in the data.

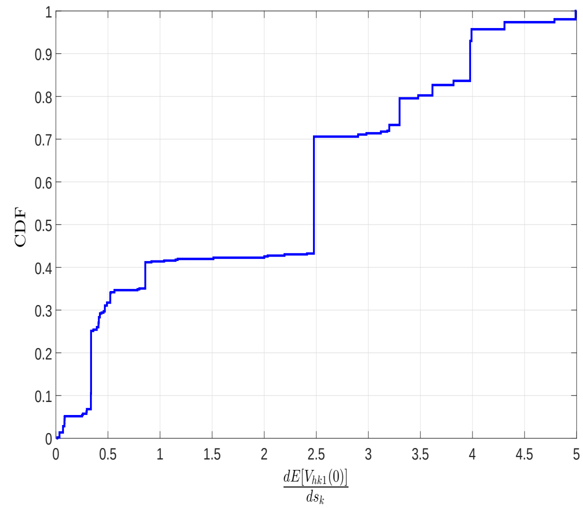
To demonstrate the implications of the estimated type distribution, Figure 5 shows the distribution of willingness to pay to increase the usage allowance by one GB on the first day of the billing cycle, $\frac{dE[V_{hk1}(0)]}{d\overline{C}_k}$ (in panel (a)), and speed by one Mb/s for the entire billing cycle, $\frac{dE[V_{hk1}(0)]}{ds_k}$, (in panel (b)). Figure 5(a) shows 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 \$0.36 (\$0.09), and the distribution is left-skewed with a small number of subscribers who are willing to pay substantial amounts. Figure 5(b) shows there is substantial variation in the preference for speed across consumers. The willingness to pay to improve speed by one Mb/s ranges from nearly zero to just over \$5, the average is \$2.02 and the median is \$2.48.

To further visualize what our estimates imply for demand, we consider subscriber behavior

Figure 5: *Distribution of Value of Increasing Usage Allowance by 1 GB and Speed by 1 Mb/s*



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



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

Note: Figure 5(a) and Figure 5(b) show the distribution of willingness to pay to increase usage allowance by 1 GB and to increase speed by 1 Mb/s, respectively.

under a linear tariff. Suppose the ISP eliminates access fees and instead sets a price p per GB, and offers just one download speed s. Because there is no fixed fee, every subscriber type consumes something under this plan. 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 calculate expected demand for five different speeds: (1) 2 Mb/s, a relatively slow speed by most standards; (2) 14.68 Mb/s, the average speed for subscribers in our data; (3) 50 Mb/s, a mid-tier speed offered currently; (4) 100 Mb/s, a top-tier speed offered currently; and (5) 1,024 Mb/s, the highest speed currently offered in North America. We present expected demand for prices of 0 to \$5 per GB.

For the average speed in our data, subscribers facing a zero price would consume an average of 2.20 GB per day, or roughly 66 GB per month. This usage is significantly reduced as the price increases. At a price of just \$1 per GB usage is cut by half, and at \$5 per GB usage is more than 10 times lower. The impact of speed can be seen by comparisons across columns. At a price of zero, expected usage at a speed of 2 Mb/s is more than 60% lower than usage at the

Table 4: *Expected Daily Usage Under a Linear Tariff*

	Expected Daily Usage (GBs)				
Price ($)	**2 Mb/s**	**14.68 Mb/s**	**50 Mb/s**	**100 Mb/s**	**1,024 Mb/s**
0.00	0.97	2.20	2.97	3.42	4.62
1.00	0.50	1.14	1.50	1.70	2.31
2.00	0.29	0.66	0.86	0.96	1.24
3.00	0.18	0.42	0.54	0.59	0.74
4.00	0.12	0.29	0.36	0.39	0.48
5.00	0.09	0.21	0.25	0.28	0.33

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

average speed in our data, while at speed of 1,024 expected usage more than doubles. Demand is much less price sensitive at lower speeds, since waiting costs form a much greater part of the subscriber's overall costs from consuming content, reducing the effect of price.

The relatively modest increase in usage with large increases in speed is intuitive and provides a reality check on our estimates. The applications used by the average consumer in 2012 do not require speeds that are substantially above the average speed in the data. Therefore, the significantly higher speeds do not induce much higher usage on average. However, even with 2012 applications these estimates demonstrate that current usage is limited by offered connection speeds. As the average user adopts more bandwidth-intensive applications, the volume of Internet traffic is likely to increase following investment in high-speed next-generation networks.

6 Implications for Managing Internet Traffic

To illustrate the implications of our estimates, we conduct several exercises to study different solutions proposed to managing the growth in Internet traffic.

6.1 The Impact of Usage-Based Pricing

We consider the impact of usage-based pricing (UBP) on usage, subscriber welfare, and the ISP's costs and revenues by comparing behavior under UBP, to consumers' plan and usage choices predicted by the demand model in various counterfactual settings. The analysis is not an equilibrium analysis, since we do not solve for the optimal offering of plans in the counterfactual setting. Instead, we use the model to simulate consumer behavior in a variety of settings.

The results are presented in Table 5. In column (1) we present the results when consumers face the UBP plans in the data, where we use the model to allocate the consumers who were on unlimited grandfathered plans to UBP plans (and calculate their usage). In column (2) we

Table 5: *Usage-based Pricing versus Unlimited Plans*

	(1)	(2)	(3)	(4)
Scenario Description				
UBP/Unlimited	UBP	Unlim	Unlim	Unlim
Plan attributes	current	current	typical US	rev-max F_k
Usage and Surplus				
Usage (GBs)	48.2	60.2	62.0	65.4
	(0.203)	(0.261)	(0.264)	(0.322)
Speed (Mb/s)	14.2	10.3	10.8	12.6
	(0.021)	(0.010)	(0.018)	(0.069)
Consumer Surplus ($)	84.7	111.9	113.5	97.1
	(0.810)	(0.791)	(0.789)	(0.810)
Revenue ($)	69.4	42.1	44.8	64.3
	(0.132)	(0.044)	(0.068)	(0.209)

Note: This table presents estimates of usage, surplus and revenue information for several scenarios. Standard errors, in parentheses, are calculated using a block-resampling methodology as described in the text.

examine the case when consumers face the same set of plans except that allowances are unlimited (and monthly fees are held constant). Not surprisingly, since the marginal price of usage is zero and monthly fees are unchanged, usage increases substantially as does consumer welfare. The ISP's revenue decreases mainly because consumers select cheaper plans that previously had a lower allowance but now differ only in speed. The sum of consumer surplus and revenue is slightly lower than column (1), and since costs are higher (due to increased usage), total surplus is lower. Relative to identical unlimited plans, UBP increases total welfare, but mainly shifts surplus from consumers to the ISP.

In column (2) we hold constant the prices of plans, the speeds, and the number of plans. It is therefore likely that we overstate the surplus lost by subscribers and the revenue gains to the ISP from UBP. We explore two ways to relax this assumption. In column (3) we present results when consumers face a typical set of unlimited plans offered by a US cable provider in 2012.[14] Interestingly, the results are very similar to those from column (2).

Next, in column (4), we return to the set of plans we see in the data but calculate the fees that maximize revenue associated with unlimited service. The ISP significantly raises fixed fees, collecting an average fee of $64.3, excluding about 7% of those who would subscribe under UBP plans. The effect is as expected, consumer surplus decreases relative to column (2) due to the higher fees, and ISP revenues increase. As in the case of column (2), UBP shifts surplus to

[14] In particular, the monthly fixed fees for the plans are $34.99, $47.99, $59.99 and $79.99 with speeds of 8, 12, 15 and 18 Mb/s.

the ISP. However, unlike the results of column (2), where the direction of the effect relative to UBP is known in advance, here consumer surplus could be higher or lower when UBP is introduced. Unlimited plans have lower marginal prices but the fixed fees are likely higher, and more importantly could change the speed of the chosen plan and therefore impact usage. The results in Table (5) show that on net consumer surplus is lower with UBP.[15] The calculation in column (4) might be over estimating the ability of the ISP to raise prices because we do not allow for entry or for a competitive response. To check this we explored introducing a DSL option, which is slower than the plans in the data, or introducing a telecomm with a fiber-to-the-node technology (such as ATT's U-verse). The results (not presented in the table) only strengthen the patterns we observe: UBP generally shifts surplus from consumers to the ISP.

6.2 Economic Viability of Next-Generation Networks

We now evaluate the economic viability of 1 Gb/s networks. This type of speed can be provided by next-generation high-speed broadband networks, such as FTTP, which has been proposed by, among others, Google.[16] DOCSIS 3.1 for cable-broadband networks is expected to be capable of similar performance.

As in the previous section, our analysis is not an equilibrium analysis: we do not have a supply model to determine the price of the FTTP plan. Furthermore, the speed is a significant extrapolation from what we see in the sample, and we (explicitly) hold the availability of content constant. One could imagine that if fast speeds become widely available, content and consumers' willingness to pay for speed will change accordingly. This increase would make our estimates of the willingness to pay for FTTP an increasingly conservative lower bound.

The results are presented in Table 6. In column (1) we display adoption and usage when FTTP is offered for free. This provides a benchmark of consumer surplus generated by the availability of a fast connection. The higher speed does indeed generate substantial surplus. However, due to a declining marginal value of speed implied by our utility function, speeds of more than 10 times those offered by the typical cable plans, imply only 2 times the surplus.

In column (2) we evaluate usage and adoption when a fee of \$70 is charged, which is what Google charges in Kansas City for Google Fiber. At this fee, but with no alternatives, over 95% of households in our population, of broadband users, are predicted to adopt FTTP and as a result usage decreases only slightly relative to a fee of zero. The next four columns examine the effect

[15] In a working paper version of this paper we provided the consumer welfare numbers by types and showed that some consumers were better off under UBP. Indeed, the net result aggregating across consumers can be sensitive to the exact estimates.

[16] See https://fiber.google.com/about/ for current offerings and expansion plans

Table 6: *Adoption of FTTP and Usage*

	(1)	(2)	(3)	(4)	(5)	(6)
Scenario Description						
Plan attributes	$F_k = 0$	$F_k = 70$	$F_k = 70$	$F_k = 70$	rev-max F_k	rev-max F_k
Competition			KC-cable	U-verse	KC-cable	U-verse
Usage and Surplus						
Usage (GBs)	138.8	136.6	134.5	134.4	133.1	132.0
	(0.855)	(0.857)	(0.856)	(0.871)	(0.901)	(0.897)
Speed (Mb/s)	1024.0	977.9	687.0	673.0	596.4	592.8
	(0.000)	(1.481)	(3.597)	(4.022)	(3.482)	(3.461)
Consumer Surplus ($)	279.4	212.9	213.2	215.5	194.3	175.0
	(1.025)	(1.014)	(0.968)	(0.981)	(0.922)	(0.889)
Revenue ($)	0.00	66.8	55.3	58.5	77.7	95.3
	(0.000)	(0.101)	(0.125)	(0.133)	(0.197)	(0.231)
FTTP Share (%)	100.0	95.5	64.7	67.1	57.2	57.2
	(0.000)	(0.145)	(0.359)	(0.397)	(0.348)	(0.351)

Note: This table presents estimates of average usage, speed consumer surplus, revenue and adoption for pricing options of FTTP as well as other broadband offerings. The adoption rate are for the population we estimated, namely, the users of broadband subscribers. Standard errors, in parentheses, are calculated using a block-resampling methodology as described in the text.

of alternative broadband offerings and different prices. In columns (3) and (4), the fee for FTTP is still $70, but either cable or U-verse options are introduced.[17] Adoption rates of FTTP fall significantly to roughly two thirds, with usage and speed being lower because some consumers choose the alternative plans. In the last two columns, when the FTTP provider charges revenue-maximizing fees, adoption of FTTP, usage and speeds fall even further. The optimal FTTP fee in Columns (5) and (6) are $106.4 and $134.4, respectively.

To address how long it will take to return the cost of investment in FTTP, we draw on estimates of the capital costs associated with the fiber network built in Kansas City by Google Fiber (Kirjner and Parameswaran (2013)).[18] If we compare FTTP to no availability of broadband at all, then from a social welfare point of view, using the estimates from column (1) suggests that these capital costs can be recovered in approximately 12 months ($3,284/$279.4). If, alternatively, broadband is offered with typical cable options, column (3) of Table 5, then the social costs are recovered in approximately 27 months ($3,284/($279.4-$113.5-$44.8)). From the ISP's

[17]The cable offerings are similar to those available in Kansas City, prices of $29.99, $39.99, $49.99, $59.99, at speeds of 3, 25, 50, 105 Mb/s. U-verse offering are $29.95, $34.95, $44.95, $64.95 for speeds of 3, 6, 18, 45.

[18]The authors estimate that it will cost $84 million dollars to "pass" 149,000 homes, or approximately $564 per household. To actually connect each home, the authors estimate it will cost 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.

perspective, the time to recover the capital costs is much higher. For example, for an existing cable service it will take approximately 149 months (\$3,284/(\$66.8 - \$44.8)) to recover the investment cost if the revenue from FTTP, priced at \$70, is compared to the revenue of a typical US cable plan (column (3) of Table 5). The exact number is sensitive to what we assume about competition before and after FTTP is introduced and whether we consider a new entrant or an existing ISP. However, the general point – that there is a large gap between social and private incentives to invest – is quite robust.

7 Conclusion

We estimate demand for residential broadband service using plan choices and high-frequency usage data. The three-part tariff plans make the usage problem dynamic and generate variation in the shadow price of usage. We show that consumers respond to this variation. Next we use the variation in shadow price to estimate a dynamic choice model and then use the estimates to evaluate the usage and welfare implications of two alternatives proposed to dealing with growth in Internet usage: 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. We find that FTTP generates significant consumer surplus but that there is 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.

There are several issues that our model does not address, and that we leave for future research. First, 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. Second, our analysis, because of data limitations, focused on GBs but not on the type of content viewed. In future work we hope to have more detailed data on the actual content, which coupled with information on TV viewing would let us explore several questions on how consumers actually use broadband. Finally, our analysis aggregated to the daily level. However, as we noted, usage is significantly higher during peak periods. This suggests that a natural way to deal with congestion is to introduce peak-load pricing. To be effective, this will require changes in popular applications, such as Netflix. We leave it to future work to explore substitution between peak and non-peak periods, which is key to the effectiveness of peak-load pricing.

REFERENCES

Abaluck, Jason and Jonathan Gruber. 2012. "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program", *American Economic Review*, 101(4): 1180-1210.

Ackerberg, Daniel. 2009. "A New Use 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.

De Fontenay, Alain, Mary Shugard, and David Sibley. 1990. "Telecommunications Demand Modelling: An Integrated View", *North-Holland*.

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

Economides, Nicholas, Katja Seim, and Brian Viard. 2008. "Quantifying the benefits of entry into local Phone Service," *The Rand Journal of Economics*, 39(3), 699.

Edell, Richard and Pravin Varaiya. 2002. "Providing Internet Access: What We Learn from IN-

DEX", Washington, DC, Editors: James Alleman and Robert Crandall, Volume: "Broadband: Should We Regulate High-Speed Internet Access?", *Brookings Institution*.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. forthcoming. "The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D". *Quarterly Journal of Economics*.

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. 2015. "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock", *American Economic Review*, 105(1), 234-271.

Grubb, Michael. 2015. "Consumer Inattention and Bill-Shock Regulation", *Review of Economic Studies*, 82(1), 219-257.

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", *American Economic Review* 103(7), 2643-2682.

Hartmann, Wesley R. 2006. "Intertemporal Effects of Consumption and Their Implications for Demand Elasticity Estimates," *Quantitative Marketing and Economics* 4, 325-49.

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.

Ketcham, Jonathan D., Claudio Lucarelli, Eugenio J. Miravete, and M. Christopher Roebuck. 2012. "Sinking, Swimming, or Learning to Swim in Medicare Part D." *American Economic*

Review, 102(6): 2639-73

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.

Miravete, Eugenio. 2003. "Choosing the Wrong Calling Plan? Ignorance and Learning", *American Economic Review*, 93(1), 297-310.

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.

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.

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.

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.

Appendix

In this Appendix, we present further analysis and results that were not included in the text due to space considerations. First, we discuss the model: we present analysis to motivate the modeling assumptions used in the paper and discuss ways to enrich the model. Second, we provide greater detail regarding the estimation and further analysis of identification. Finally, we present additional results including descriptive evidence on the importance of consumer heterogeneity and details on the estimated type distribution.

1 Extensions of the Model

Day-of-Week Dependence

If a consumer's online habits vary by the day of week, the distribution of the preference shocks should include a day-of-week component (which would violate the iid assumption). In principle, this can easily be dealt with by adding an additional state variable that captures the day of the week. Table 7 presents average daily usage by day of week. We find no consistent and statistically-significant difference across days. To check that the lack of day-specific demand is not driven by some odd behavior in our data we also look at two other operators for roughly the same time period. The results for these operators also exhibit little difference in the level of activity across days, suggesting that it is not necessary to add day of week to the state vector.[19]

Table 7: *Average Daily Usage by Day of Week*

Day of Week	**Daily Usage (GBs)**
Sunday	1.55
Monday	1.59
Tuesday	1.50
Wednesday	1.47
Thursday	1.46
Friday	1.46
Saturday	1.48

Note: This table presents average daily usage during the May-June 2012 billing cycle.

Transferability of Content Across Days

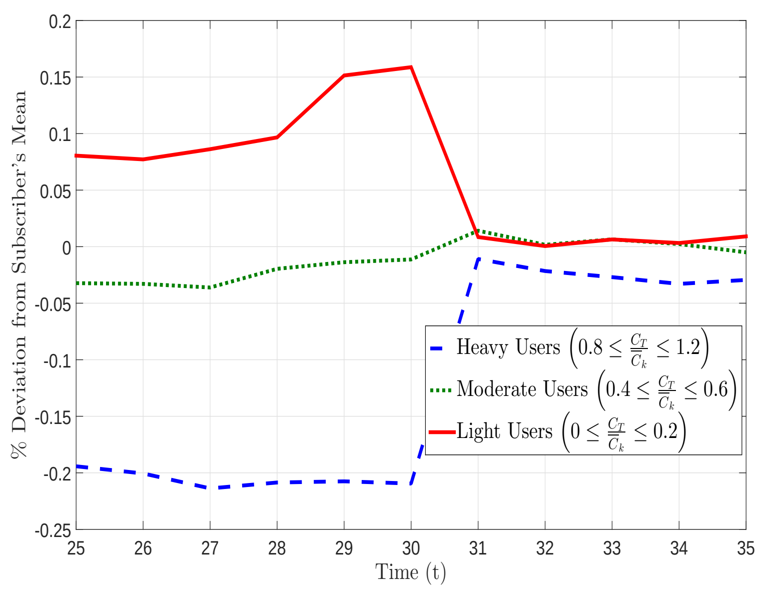
Another violation of the iid assumptions would occur if subscribers postpone consumption from one day to another. The end of one billing cycle, and the beginning of the next, presents

[19]Note that any variation by day of week is separately identified from patterns in usage across the month due to usage-based pricing, since consumers have staggered billing cycles.

the best opportunity to find evidence that subscribers have postponed consumption of content until a later time. Content is most likely to be postponed when a subscriber knows that the usage allowance is refreshed, and they don't have to postpone consumption by more than a few days. Thus, if we find little evidence of content being transferred at the end of the billing cycle in these situations, it suggests that transfer of content across days is unlikely to occur elsewhere.

To look for evidence of content being transferred across billing cycles, we enrich the end-of-month analysis from the main text. First, for each day and every consumer, we calculate the percentage deviation in daily usage from the consumer's own mean. We then classify consumers into groups based upon their cumulative usage at the end of the billing cycle: light $\left(0 \leq \frac{C_T}{\overline{C}_k} \leq 0.2\right)$, moderate $\left(0.4 \leq \frac{C_T}{\overline{C}_k} \leq 0.6\right)$, and heavy $\left(0.8 \leq \frac{C_T}{\overline{C}_k} \leq 1.2\right)$. For each of these groups, we calculate the average percentage deviation over each of the last five days of the billing cycle and the first five days of the next billing cycle. Figure 6 presents the results of these calculations. If content was being transferred to the next billing cycle we would expect to see a higher than average usage on the first few days, especially for the heavy users. We observe no evidence of content being transferred across billing cycles for those consumers near the allowance.

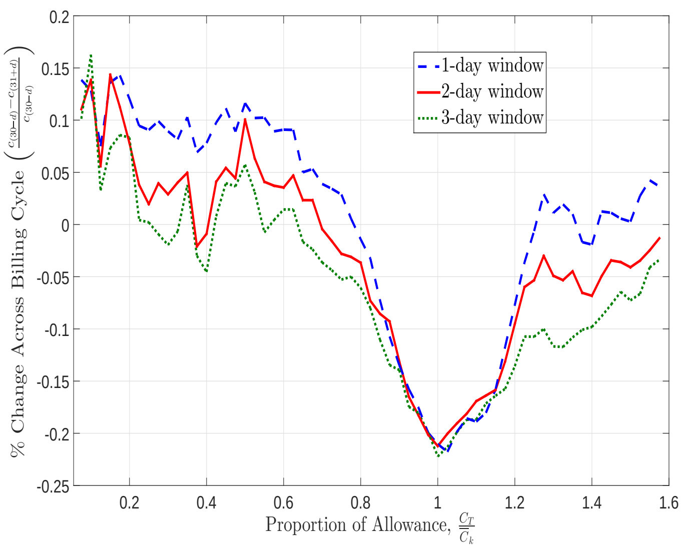
Figure 6: *Transferability of Content, Across-Month Dynamics*



Note: This figure presents the average percentage deviation from a subscriber's own daily average for three groups of subscribers on usage-based plans (light, moderate and heavy) for the last 5 days of the May-June 2012 billing cycle and the first 5 days of the next billing cycle.

To further demonstrate the robustness of our results presented in the main text, Figure 7 presents the results from a calculation identical to that used to generate Figure 2, except we perform the calculation using a 1-day, 2-day, and 3-day window. Specifically, the 1-day window is identical to the results presented in the main text. The results for the 2-day (3-day) window are similar, but rather than the difference in usage on the first and last days of the billing cycle being used in the calculation, average usage on the last two (three) and first two (three) days of the billing cycle are used. The results are similar with the only noticeable difference occurring for those subscribers well over the allowance by the end of the month. However, this variation is largely due to the very small number of subscribers who substantially exceed their allowance.

Figure 7: *Across-Month Dynamics*



Note: This figure presents how the percentage difference between average usage during the last days of a billing cycle and average usage during the first days of the next varies with the proportion of the allowance consumed by a subscriber at the end of the billing cycle. The figure presents results using a one, two, and three day window for calculating the averages.

Serial Correlation

To examine the possibility of serial correlation in the realizations of the preference shock we cannot simply look at serial correlation in usage since our model predicts that (positive) serial correlation will arise through the shadow price faced by a subscriber. We therefore examine whether serial correlation in usage varies with how close a consumer is to the allowance. Serial correlation due to correlation in the shadow price predicts that closeness to the allowance will impact serial correlation. Under the alternative, of serial correlation in the preference shock, this will not be the case.[20]

[20]Note that the across-month analysis largely rules out negative serial correlation, or mean reversion, as the explanation for our results, so we do not discuss it further here.

Table 8: *Identifying Source of Serial Correlation*

	(1)	**(2)**
$\frac{C_{jT}}{C_{jk}}$	0.047**	0.077**
	(0.002)	(0.003)
$\left(\frac{C_{jT}}{C_{jk}}\right)^2$	–	-0.004**
	–	(0.001)
Constant	0.151**	0.139**
	(0.001)	(0.002)
Observations	42,485	42,485

Note: This table presents OLS estimates from regressing the correlation coefficient for each subscriber (ρ_j from the regression, $\widetilde{c}_{jt} = \rho_j \widetilde{c}_{j(t-1)} + \eta_{jt}$) on the fraction of the allowance used by a subscriber, $\frac{C_{jT}}{C_{jk}}$. The regression uses 42,485 observations. Asterisks denote statistical significance: ** 1% level, * 5% level.

Specifically, we calculate the deviation in daily usage from the subscriber's mean as $\widetilde{c}_{jt} = c_{jt} - \frac{1}{T}\sum_{j=1}^{T} c_{jt}$, for each subscriber on each day. We then calculate the measure of serial correlation for each subscriber, ρ_j, by regressing these deviations on their own lags ($\widetilde{c}_{jt} = \rho_j \widetilde{c}_{j(t-1)} + \eta_{jt}$). The median correlation is 0.151 and average correlation coefficient is 0.176.

To identify those consumers most likely to exhibit positive serial correlation in their usage we regress the estimate of ρ_j on the fraction of the allowance used by the subscriber at the end of the billing cycle, $\frac{C_{jT}}{C_{jk}}$. The results are presented in the first column of Table 8. We find that consumers near their allowance by the end of a billing cycle exhibit more serial correlation compared to those who use a small portion of their allowance. The second column of Table 8 presents the results when we add a quadratic term to allow for nonlinearities in the relationship. We also find that the quadratic term has a negative sign, which suggests a diminishing effect, such that those well over the allowance exhibit less serial correlation. While this effect is statistically significant, it is quite small in magnitude.

Lumpiness of Usage

It can be argued that Internet consumption is increasingly "lumpy" as a larger proportion of traffic consists of either bandwidth-intensive cloud-based services (e.g. iTunes, Dropbox, etc) or online video offerings (e.g., Netflix, YouTube, etc). For these types of content, consumption often involves a single download or upload of substantial size. One could argue that a continuous distribution like the log-normal cannot replicate the usage patterns in the data because of the lumpiness. However, as we show in Figure 4, in the main text, the model estimates successfully replicate the variation from day to day in a subscriber's usage.

Additional Price Sensitivity

There are at least a couple of other ways we could add additional price sensitivity into the model. First, the marginal, or shadow, price of usage in the model is given by

$$\widetilde{p}_k(c_t, C_{t-1}) = \begin{cases} p_k & \text{if } \mathcal{O}_{tk}(c_t) > 0 \\ \rho_1 \frac{dE[V_{hk(t+1}(C_{t-1}+c_t)]}{dc_t} & \text{if } \mathcal{O}_{tk}(c_t) = 0 \ . \end{cases}$$

We restrict ρ_1 to equal one for the analysis in the main text. Namely, at each point in the billing cycle, either the subscriber is making marginal decisions on usage facing the overage price, p_k, or fully internalizing the impact of current usage decisions on the possibility of overages, such that the perceived price equals $\frac{dE[V_{hk(t+1}(C_{t-1}+c_t)]}{dc_t}$. By allowing ρ_1 to differ from 1 we allow for consumers who do not fully internalize the impact of current usage until overages are actually incurred, or overreact to the possibility of overages.

We estimate ρ_1 in a way similar to the other five parameters by using three points of support for ρ_1 (0.5, 1, and 1.5).[21] These points of support for ρ_1 permit for subscribers who (i) respond more to actual overages than to changes in the shadow price ($\rho_1 = 0.5$), (ii) fully internalize the impact of usage on the possibility of overages ($\rho_1 = 1$), and (iii) overreact to the possibility of incurring overage charges ($\rho_1 = 1.5$). The parameter ρ_1 is identified by behavior just before and just after the allowance is exceeded. If there is no change right around the allowance then usage is consistent with $\rho_1 = 1$. On the other hand, a difference in behavior just before and just after the allowance is reached is consistent with ρ_1 different from 1.

Our estimate of the marginal distribution of ρ_1 provides support for the assumption in the main text. Specifically, we estimate that 91.1% of subscribers have ρ_1 equal to one, while 3.1% equal 0.5 and 5.9% equal 1.5. Most importantly, we find that this small number of subscribers with ρ_1 different from one has little implication for our counterfactual results.

An alternative way to introduce additional price sensitivity is by adding a coefficient in front of the last term in the utility from content: the utility from the numeraire. In the text, we normalize the marginal utility from y_t to 1, which means that, as we discuss above, the price sensitivity is captured by β_h. A more general model would multiply y_t by a coefficient. We did not explore estimating this coefficient, as it is not clear that is identified in our data.

[21] We only use usage-based plans for this estimation due to the absence of overages for the grandfathered unlimited plans.

Plan Choice Model

Our model of plan choice assumes that consumers choose the plan that maximizes the dynamic value and that there are no errors in plan selection. We can relax this assumption, by assuming that each type of subscriber maximizes $\rho_2 E\left[V_{hk0}(0)\right] - F_k + \epsilon_{hk}$, where ϵ_{hk} are *iid* extreme value shocks and ρ_2 is a parameter that scales the error shock and the expected utility from the plan.

We could also extend the plan choice model by letting the utility that the subscriber maximizes be $E\left[V_{hk0}(0)\right] - \rho_3 F_k$, (or $\rho_2 E\left[V_{hk0}(0)\right] - \rho_3 F_k + \epsilon_{hk}$ if we want to include the extension of the previous paragraph). Note that ρ_3, the price coefficient will pick up the relative importance of price of the plan and the expected value from usage. For a fully rational consumer $\rho_3 = 1$.

2 Econometric Details

2.1 Step 1: Solving the Model

As we describe in the main text, in the first step of the estimation algorithm we solve the dynamic problem for a large number of types, once for each type, and store the optimal policy.

For a plan, k, and subscriber type, h, we solve the finite-horizon dynamic program recursively. 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 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 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,

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

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.

2.2 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.

As we describe in the main text, our estimates of the weights are chosen to maximize the objective function. The weighting matrix, $\widehat{\mathbf{V}}^{-1}$, would ideally be the variance covariance matrix of $\widehat{\mathbf{m}}_k^{dat}$, ensuring that more variable moments receive less weight. This choice of weighting matrix in our application is problematic. Specifically, since our approach relies on state-specific moments, which are aggregated across a large number of types, the variance of the moments can be quite small. These very low variance moments cause numerical instability during the optimization of the objective function. For this reason, we set $\widehat{\mathbf{V}}^{-1}$ equal to the identity matrix.

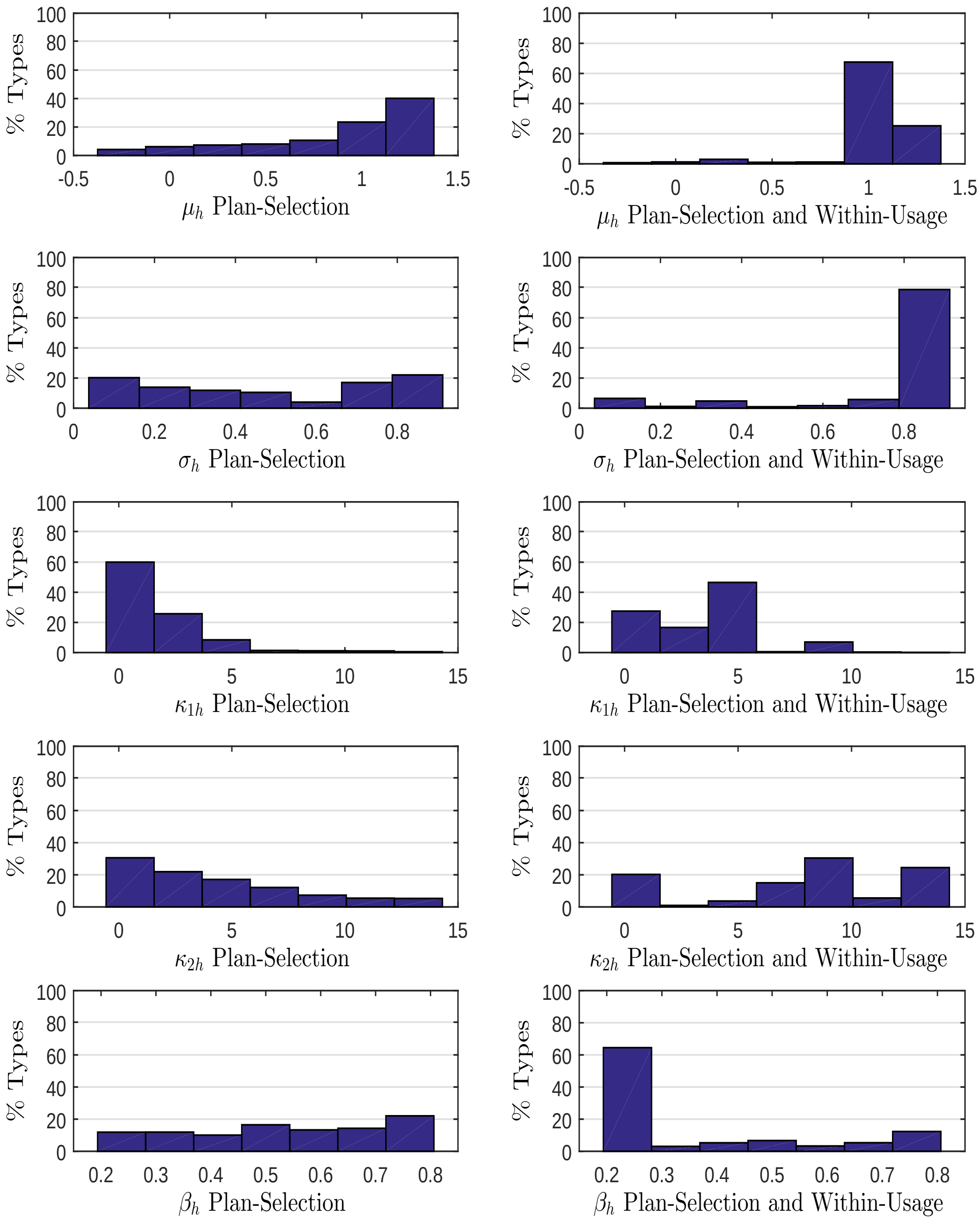
The richness of the data, along with the low dimensionality of the state space, (C_{t-1}, t), allows a flexible approach for recovering moments from the data to match with the model.

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 $\widehat{\Gamma}_{kt}(C) \in [0,1]$, ensuring that we fit the tails of the usage distribution. We use a normal kernel with an adaptive bandwidth to smooth the empirical cdf.

Figure 8: *Sources of Identification: Plan Selection and Usage*



Note: This figure presents the marginal distribution of each parameter, when only information on optimal plan selection is used and uniform weights are applied, and when the weights are chosen using information on optimal plan and to match usage moments from the data.

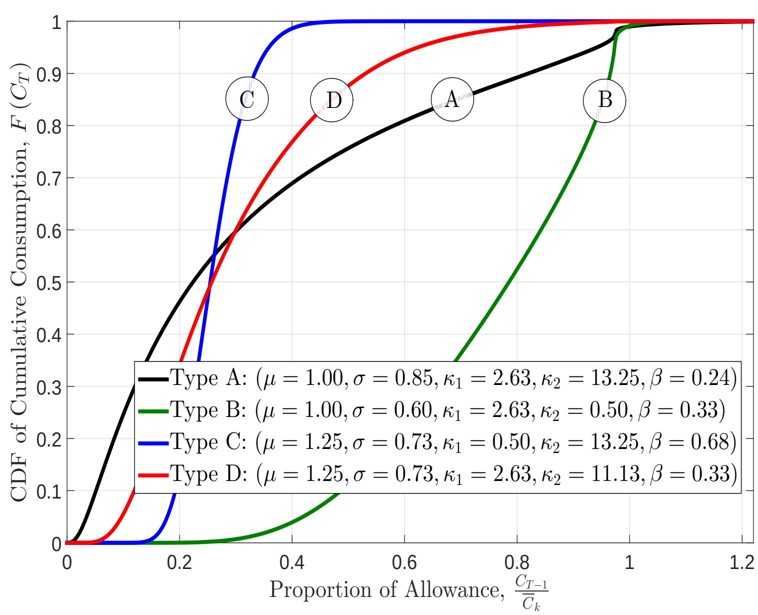
We recover the moments of usage at each state by estimating a smooth surface using a nearest-neighbor approach. Consider a point in the state space, (C_{t-1}, t). A neighbor is an observation in the data for which the subscriber is t days into the billing cycle and cumulative consumption up until day 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, we multiply $\widehat{E}\left[c_{kt}^*(C_{t-1})\right]$ 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 of the 8 plans, or $8 * 120,000 = 960,000$ moments in total.

Figure 9: *Predicted Behavior by Type, CDF of* C_T

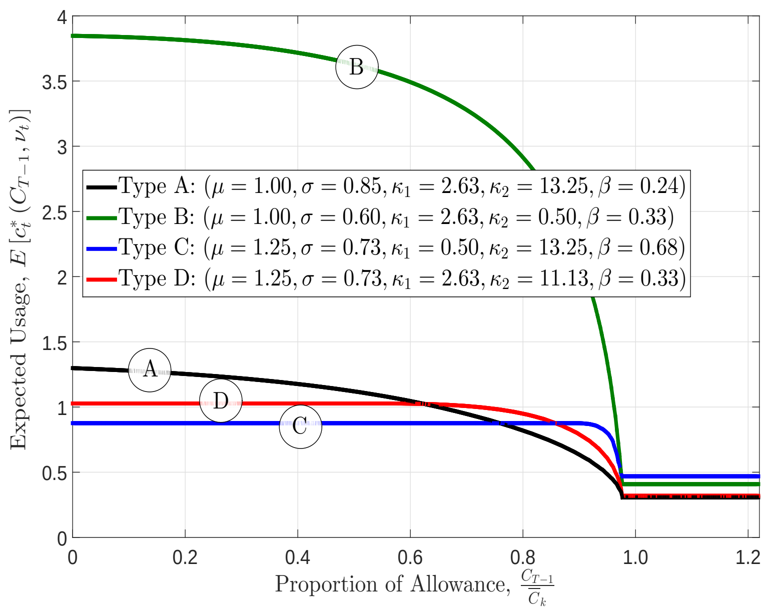


Note: This figure presents the cdf of cumulative consumption on the last day of the billing cycle, C_T, for the types that received the greatest estimated weight on one of the usage-based plans in our data.

2.3 Identification: Plan Selection and Usage

In this subsection we provide additional results to demonstrate the relative importance of plan selection and usage. In particular, in Figure 8 we present the marginal distribution of all five parameters using only plan selection, in the left graphs, and both usage and plan selection in the right graphs. The results confirm that usage information is driving much of the results.

Figure 10: *Predicted Behavior by Type: Expected Usage* $E\left[c_T^*\left(C_{T-1},\nu_t\right)\right]$



Note: This figure presents expected usage on the last day of the billing cycle, $c_T^*\left(C_{T-1}\right)$, for the types that received the greatest estimated weight on one of the usage-based plans in our data.

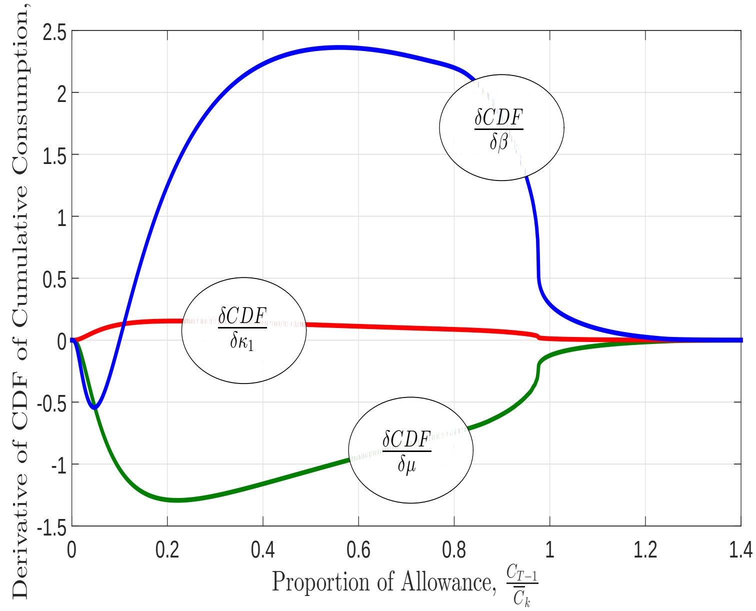
2.4 Identification: Variation in Behavior Across Types

To demonstrate the differences in behavior between types we look at the heterogeneity in behaviors within a particular plan. In Figures 9 and 10, we plot the model's predicted behavior for the four types with the greatest estimated weights on a particular usage-based plan. Figure 9 plots the cdf of cumulative consumption on the final day of the billing cycle, C_T, for each of the types. Figure 10 plots expected usage on the last day of the billing cycle, conditional on each possible level of cumulative consumption on the previous days, C_{T-1}.[22] The functions in this Figure highlight price sensitivity, through the change in expected usage as the fraction of the allowance used, $\frac{C_{T-1}}{\overline{C}_k}$, increases. This price sensitivity also reveals information about variation in usage. A subscriber whose cumulative usage is well below the allowance, but whose expected

[22] In estimation, we weight this moment by the probability of reaching each of these states to preserve linearity.

usage changes with small movements in the fraction of the allowance used, has some chance of very high usage which will vary depending upon how many GB are subject to overages.

Figure 11: *Perturbation of Parameters, CDF of* C_T



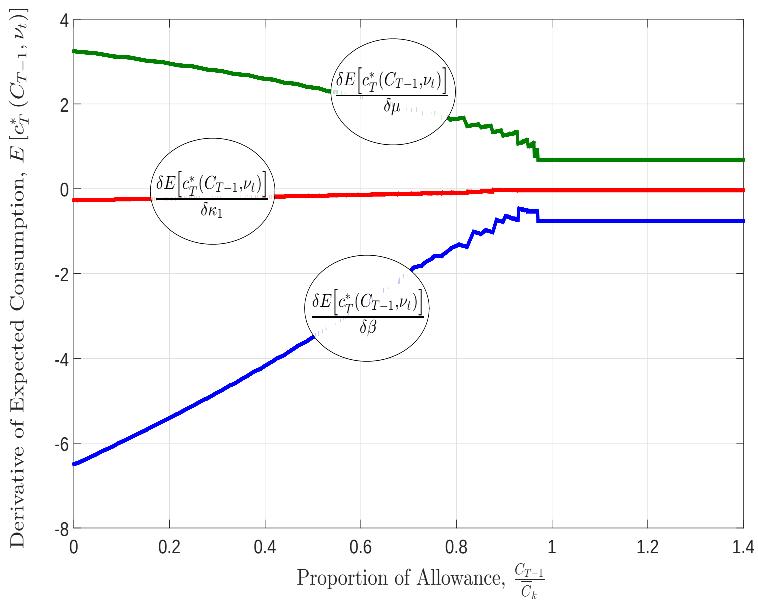
Note: This figure presents the derivative of the cdf of cumulative consumption on the last day of the billing cycle, C_T, with respect to each of the type-specific parameters, μ, κ_1, and β, for the most common type, $(\mu = 1.00, \sigma = 0.60, \kappa_1 = 2.63, \kappa_2 = 0.50, \beta = 0.33)$.

Figures 9 and 10 make clear that variation in the parameters induces substantially different behaviors even within a group of types that prefer the same plan. For example, type A, with parameters $(\mu = 1.00, \sigma = 0.85, \kappa_1 = 2.63, \kappa_2 = 13.25, \beta = 0.24)$, typically has low usage with occasionally very high usage (think of this user as one that will occasionally watch online video). Type B, with parameters $(\mu = 1.00, \sigma = 0.60, \kappa_1 = 2.63, \kappa_2 = 0.5, \beta = 0.33)$, has typical high usage, a greater probability of reaching high cumulative consumption states, and a quite high usage elasticity. This type can be considered as someone who regularly watches online video. In contrast, type C, characterized by the the vector $(\mu = 1.25, \sigma = 0.73, \kappa_1 = 0.50, \kappa_2 = 13.25, \beta = 0.68)$ has a very low probability of reaching high cumulative consumption states and a relatively low usage elasticity. This type seems to mainly use the broadband connection for applications that are less data intensive, such as web surfing, and seems to choose this plan for the greater speed.

To further isolate the role that each parameter has in determining type-specific behavior, and the moments used in estimation, we consider how perturbations of the parameter vector are reflected in the particular moments we consider. We consider the most common type in our data, $(\mu = 1.00, \sigma = 0.60, \kappa_1 = 2.63, \kappa_2 = 0.50, \beta = 0.33)$, accounting for over 28% of the population.

Figure 11 presents the derivative of the cdf of cumulative consumption on the last day of the billing cycle, C_T, with respect to μ, κ_1, and β. Figure 12 presents the derivative of optimal consumption on the last day of the billing cycle, $c_T^*(C_{T-1})$, for each possible level of cumulative consumption, C_{T-1}.

Figure 12: *Perturbation of Parameters, Expected Usage* $c_T^*(C_{T-1})$



Note: This figure presents the derivative of expected consumption conditional on reaching a state on the last day of the billing cycle, $c_T^*(C_{T-1})$, with respect to each of the type-specific parameters, μ, κ_1, and β, for the most common type, $(\mu = 1.00, \sigma = 0.60, \kappa_1 = 2.63, \kappa_2 = 0.50, \beta = 0.33)$.

For each of the parameters, perturbations result in quite different behavioral responses. An increase in μ results the entire distribution of cumulative consumption shifting to the right, i.e., the cdf is lower at each point, particularly at low cumulative consumption states. Conversely, an increase in κ_1 shifts the entire distribution to the left, i.e., the cdf is greater at each point, although relatively similarly across the entire distribution. The most interesting response comes when β is increased, which results in a more concentrated distribution, i.e., the frequency of very low and very high cumulative consumption states are reduced. Figure 12 shows that perturbations of each of the parameters has a differing effect on expected consumption conditional on reaching a particular cumulative consumption state. Again, κ_1 has a moderate and similar effect on expected consumption regardless of the state. An increase in β results in decreased usage at all states while an increase in μ results in increased usage at all states. The absolute value of these changes is smallest near the usage allowance.

2.5 Grid Density

Our estimates presented in the main text form a densely-populated grid of types. Specifically, for each of the five parameters, we consider seven points of support for a total of 16,807 types. Given the available data and our approach, which searches for an optimal mixture of types to match state-specific moments from each plan, we find that we cannot increase the density of the grid without severe multicollinearity problems (i.e., types behave too similarly at all states). Yet, it is of interest to understand how sparse the grid of types can be and still adequately match the observed behavior, since in more complex settings with a higher-dimensional parameter vector, it may be impossible to solve for such a dense grid.

Table 9: *Estimated Weights for Top-20 Types, Coarser Grid*

Type	μ_h	σ_h	κ_{1h}	κ_{2h}	β_h	θ_h	$se(\theta_h)$	**Cumulative**
1	1.25	0.85	9.00	4.75	0.2375	0.223	0.002	0.223
1	1.25	0.85	4.75	4.75	0.2375	0.164	0.001	0.387
1	1.25	0.85	4.75	13.25	0.2375	0.104	0.001	0.491
1	1.25	0.85	4.75	9.00	0.4125	0.084	0.001	0.575
1	0.75	0.85	4.75	0.50	0.4125	0.044	0.001	0.619
1	1.25	0.60	4.75	0.50	0.2375	0.032	0.001	0.651
1	1.25	0.85	0.50	13.25	0.7625	0.029	0.001	0.680
1	1.25	0.60	0.50	4.75	0.7625	0.025	0.001	0.705
1	0.25	0.85	0.50	0.50	0.7625	0.024	0.001	0.729
1	1.25	0.85	0.50	9.00	0.7625	0.023	0.001	0.752
1	1.25	0.60	0.50	13.25	0.7625	0.023	0.001	0.775
1	1.25	0.85	0.50	13.25	0.4125	0.022	0.001	0.797
1	0.75	0.85	4.75	0.50	0.2375	0.021	0.001	0.818
1	0.25	0.85	9.00	13.25	0.7625	0.020	0.001	0.839
1	-0.25	0.85	4.75	0.50	0.7625	0.017	0.001	0.856
1	0.25	0.85	4.75	0.50	0.7625	0.015	0.001	0.871
1	-0.25	0.60	0.50	0.50	0.7625	0.014	0.001	0.885
1	1.25	0.85	4.75	9.00	0.2375	0.013	0.001	0.898
1	0.75	0.60	0.50	4.75	0.7625	0.013	0.001	0.910
1	0.25	0.35	0.50	0.50	0.4125	0.011	0.002	0.922

Note: The table presents the 20 types that receive the highest weight, when we estimate the model on a coarser grid. It is meant to demonstrate the sensitivity to the choice of grid.

To show that our approach performs well even when the density of the grid of types is substantially reduced, we take a straightforward approach. For each parameter, we consider the same range of support (minimum and maximum unchanged), but remove every other point of support. This leaves four points of support along each dimension for a total of 1,024 types (4^5). We then follow the same steps to estimate weights for each type. We find that 36 types get positive weight. The top-5 types account for 62% of the mass, the top-10 account for 75%, and the top-20 for 92%. We present the estimated weights for the top-20 types in Table 9.

Table 10: *Expected Daily Usage Under a Linear Tariff, Coarse Grid*

	Expected Daily Usage (GBs)				
Price ($)	**2 Mb/s**	**14.68 Mb/s**	**50 Mb/s**	**100 Mb/s**	**1,024 Mb/s**
0.00	0.45	2.22	2.99	3.37	4.28
1.00	0.27	1.26	1.74	1.97	2.57
2.00	0.19	0.80	1.06	1.18	1.48
3.00	0.14	0.54	0.70	0.77	0.94
4.00	0.11	0.39	0.49	0.53	0.63
5.00	0.09	0.29	0.36	0.38	0.45

Note: This table presents the expected daily usage averaged across all subscriber types when facing a linear tariff when we estimate the model on a coarser grid. It is meant to demonstrate the sensitivity to the choice of grid.

Despite the differences between the estimates, we find the implications of the type distribution for overall demand are rather similar. To visualize this, Table 10 presents average daily usage under the same set of linear tariffs as Table 4. The largest differences in predicted daily usage occur for the lowest speeds, while the predictions for higher speeds are quite similar. The similarity of the results gives us confidence that our approach can be applied in more complex and higher-dimensional problems, where a less-dense grid of types is necessary.

3 Additional Results

3.1 Descriptive Statistics of Consumer Heterogeneity

Consumers in our data exhibit a wide range of behaviors. To demonstrate this, we first present the substantial variation in usage across subscribers, and how it has changed over time. Second, we present the variation in the fraction of the allowance used by subscribers. Both these dimensions of usage motivate our need for a rich distribution of consumer heterogeneity.

Table 11 presents usage statistics in May 2011 and May 2012, along with the corresponding year-on-year growth rates. 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 higher for more intensive users, but in percentage terms is monotonically higher for less-intensive users.

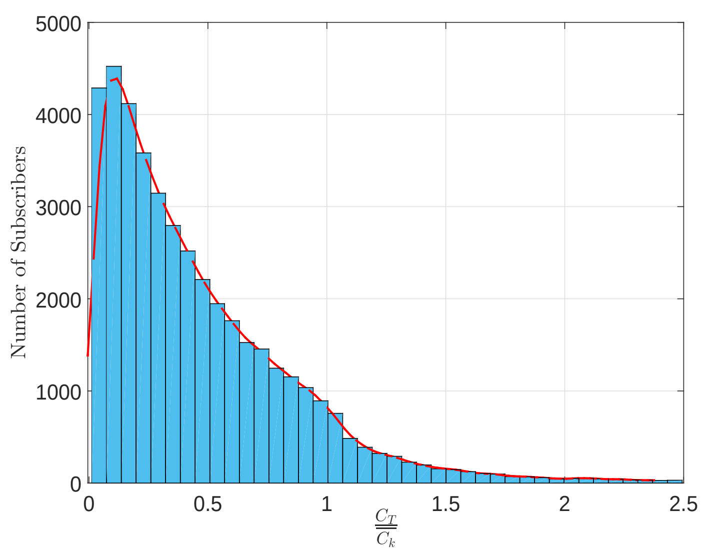
Figure 13 presents the variation in the fraction of the allowance used by consumers in June

Table 11: *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.

Figure 13: *Proportion of Allowance Used*



Note: This figure presents a histogram where each observation represents a consumer's monthly usage relative to their allowance, both in GBs.

Table 12: *Descriptive Statistics for Type Distribution*

	Min	**Max**	**Median**	**Mean**	**S.D.**
mean of shocks (μ)	-0.25	1.25	1.00	0.93	0.35
s.d. of shocks (σ)	0.10	0.85	0.85	0.77	0.18
opp cost of content (κ_1)	0.50	13.25	4.75	3.83	2.45
pref for speed (κ_2)	0.50	13.25	9.00	7.22	4.84
curvature (β)	0.24	0.76	0.24	0.38	0.21

Note: These statistics reflect the estimated distribution of types after removing those types with weights $\theta_h <$ 0.0001 and re normalizing the weights of the remaining 53 types.

2012. As mentioned in the text, 10% of subscribers exceed their allowance during this billing cycle. The average share of the allowance used is approximately 49%. For those subscribers exceeding their allowance, the median overage is 17 GBs and the corresponding charge is $51.19.

3.2 Estimated Type Distribution

Our flexible econometric approach yields a discrete distribution of types, which is difficult to describe concisely. For this reason, details were omitted from the main text, but we provide them here.

Table 12 presents descriptive statistics for the type distribution, and highlights the importance of the flexibility of our econometric approach. We find a substantial range of heterogeneity for each of the parameters characterizing a type. There are also noticeable differences in the median and means, which highlight the skewness of the distribution along each dimension. In particular, the mean for the curvature parameter, β, is more than 50% greater than the median.

To further visualize the heterogeneity among types, Table 13 presents the 20 types with the largest estimated weights, which account for 90% of the mass for the overall distribution.

Table 13: *Estimated Weights for Top-20 Types*

Type	μ_h	σ_h	κ_{1h}	κ_{2h}	β_h	θ_h	$se\,(\theta_h)$	Cumulative
1	1.000	0.850	4.750	9.000	0.238	0.281	0.002	0.281
2	1.000	0.850	4.750	0.500	0.238	0.150	0.002	0.431
3	1.000	0.850	2.625	13.250	0.238	0.095	0.001	0.526
4	1.000	0.850	0.500	13.250	0.763	0.063	0.001	0.589
5	1.250	0.850	9.000	6.875	0.238	0.063	0.001	0.652
6	1.000	0.600	2.625	0.500	0.325	0.038	0.002	0.690
7	1.250	0.725	0.500	13.250	0.675	0.027	0.001	0.717
8	1.250	0.725	2.625	11.125	0.325	0.024	0.001	0.741
9	0.250	0.850	9.000	13.250	0.763	0.022	0.001	0.763
10	1.250	0.850	0.500	13.250	0.413	0.019	0.001	0.782
11	1.000	0.850	2.625	11.125	0.238	0.016	0.001	0.798
12	-0.250	0.850	2.625	0.500	0.675	0.014	0.001	0.812
13	-0.250	0.850	2.625	6.875	0.763	0.014	0.001	0.826
14	1.250	0.350	0.500	4.750	0.675	0.013	0.001	0.839
15	0.250	0.350	0.500	0.500	0.413	0.012	0.001	0.852
16	-0.250	0.725	4.750	0.500	0.763	0.011	0.001	0.863
17	1.250	0.850	2.625	4.750	0.325	0.010	0.001	0.873
18	0.750	0.475	2.625	0.500	0.413	0.010	0.001	0.882
19	-0.250	0.725	0.500	2.625	0.675	0.008	0.001	0.890
20	0.250	0.475	0.500	0.500	0.500	0.007	0.001	0.900

Note: The table presents the 20 types that receive the highest weight. Since these types account for 90% of the mass the table essentially characterizes the type distribution.