

SUPPLEMENT TO “USAGE-BASED PRICING AND DEMAND FOR
RESIDENTIAL BROADBAND”: APPENDIX
(*Econometrica*, Vol. 84, No. 2, March 2016, 411–443)

BY AVIV NEVO, JOHN L. TURNER, AND JONATHAN W. WILLIAMS

IN THIS APPENDIX, we present further analysis and results that were not included in the text due to space considerations. In Section S.1, we discuss the model: we present further analysis to motivate the modeling assumptions used in the paper and discuss ways to enrich the model. In Section S.2, we provide greater detail regarding the estimation and further analysis of identification.

S.1. EXTENSIONS OF THE MODEL

S.1.1. *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 i.i.d. assumption). In principle, this can easily be dealt with by adding an additional state variable that captures the day of the week. Table VIII 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.¹

S.1.2. *Transferability of Content Across Days*

Another violation of the i.i.d. 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 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 will soon be refreshed, and they do not 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

¹Note 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.

TABLE VIII
AVERAGE DAILY USAGE BY DAY OF WEEK^a

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

^aThis table presents average daily usage during the May–June 2012 billing cycle.

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 ($0 \leq \frac{C_T}{C_k} \leq 0.2$), moderate ($0.4 \leq \frac{C_T}{C_k} \leq 0.6$), and heavy ($0.8 \leq \frac{C_T}{C_k} \leq 1.2$). 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 7

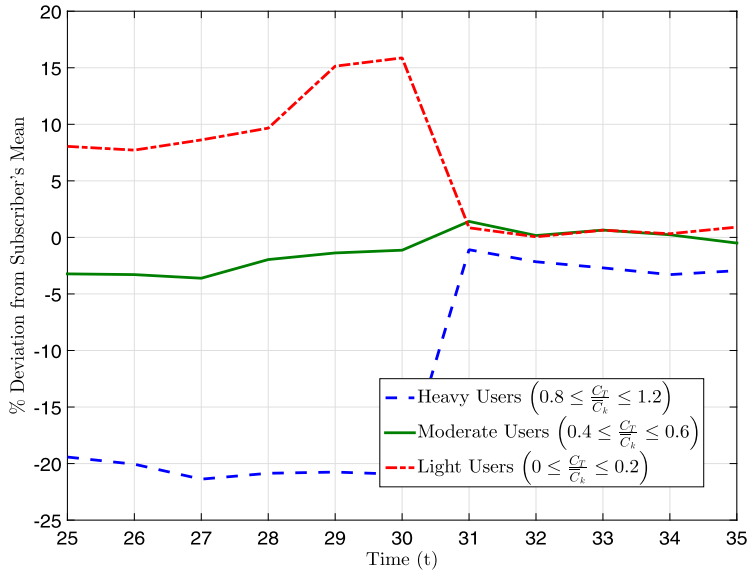


FIGURE 7.—Transferability of content, across-month dynamics. *Note:* This figure presents the average percentage deviation from a subscriber’s own daily average for three groups of subscribers on usage-based plans (light, moderate, and heavy) for the last five days of the May–June 2012 billing cycle and the first five days of the next billing cycle.

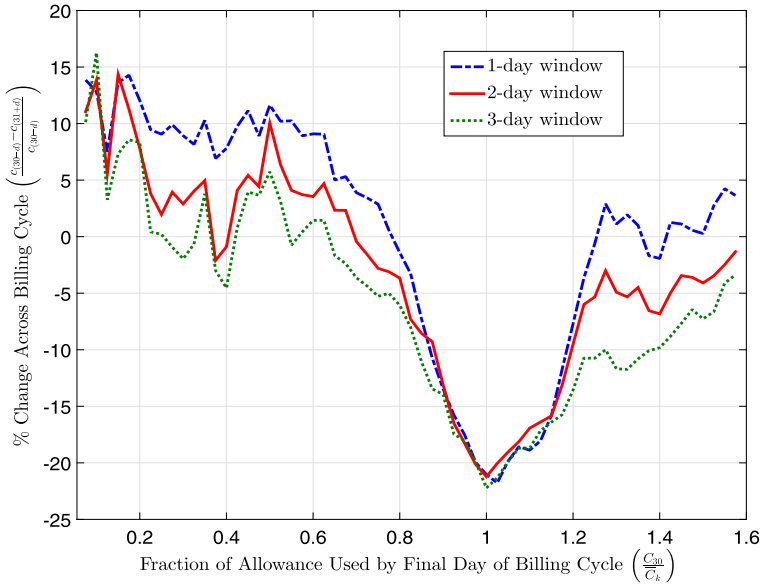


FIGURE 8.—Across-month dynamics. *Note:* This figure presents how the percentage difference between average usage during the last days of a billing cycle and average usage during the first days of the next varies with the proportion of the allowance consumed by a subscriber at the end of the billing cycle. The figure presents results using a 1-, 2-, and 3-day window for calculating the averages.

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.

To further demonstrate the robustness of our results presented in the main text, Figure 8 presents the results from a calculation identical to that used to generate Figure 3, 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 is 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.

S.1.3. *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.²

Specifically, we calculate the deviation in daily usage from the subscriber's mean as $\tilde{c}_{jt} = c_{jt} - \frac{1}{T} \sum_{j=1}^T c_{jt}$, for each subscriber on each day. We then calculate the measure of serial correlation for each subscriber, ϕ_j , by regressing these deviations on their own lags ($\tilde{c}_{jt} = \phi_j \tilde{c}_{j(t-1)} + \eta_{jt}$). The median correlation is 0.151 and the 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 ϕ_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 IX. 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 IX
IDENTIFYING SOURCE OF SERIAL CORRELATION^a

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

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

²Note 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 IX 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.

S.1.4. *Response to the Shadow Price versus Overage Charge*

In our model, rational consumers treat shadow price variation as if it is actual variation in price. We provide evidence in Section 2.2 in the main text that consumers indeed respond to the shadow price. But, in principle, they might respond differently to a change in the shadow price versus a change in the actual price, namely the overage charge. To examine this, we do two things. First, we estimate a model that allows for a differential response. Next, we estimate a model that only uses realized overage charges (and plan choice) to estimate the model.

The marginal, or shadow, price of usage in the model is given by

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

We restrict ρ to equal 1 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 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 in a way similar to the other five parameters by using three points of support for ρ (0.5, 1, and 1.5).³ These points of support for ρ permit subscribers who (i) respond more to actual overages than to changes in the shadow price ($\rho = 0.5$), (ii) fully internalize the impact of usage on the possibility of overages ($\rho = 1$), and (iii) overreact to the possibility of incurring overage charges ($\rho = 1.5$). The parameter ρ 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 $\rho = 1$. On the other hand, a difference in behavior just before and just after the allowance is reached is consistent with ρ different from 1.

³We only use usage-based plans for this estimation due to the absence of overages for the grandfathered unlimited plans.

Our estimate of the marginal distribution of ρ provides support for the assumption in the main text. Specifically, we estimate that 91.0% of subscribers have ρ equal to 1, while 3.1% equal 0.5 and 5.9% equal 1.5. Most importantly, we find that this small number of subscribers with ρ different from 1 has little implication for our counterfactual results.

We also estimate the model with the restriction that $\rho = 0$. This is equivalent to not using the variation in shadow price, and only using the variation in the actual price. We expect this to yield a flatter demand curve that is less responsive to price. The intuition is simple. Before reaching the allowance, we saw that consumers reduce their consumption. The dynamic model interprets this as a response to the shadow price, but the simpler model assumes that this is noise and just reduces the average usage at a price of zero. This intuition is confirmed from estimates of the simple model.

We estimate this model using the exact same moments as the full model, but by setting $\rho = 0$ we assume that the price is zero until the consumer reaches the allowance. In principle, there are simpler ways to estimate this model. But by using the same moments, we believe we can separate the effect of dynamics from differences in estimation methods. The estimation yields a distribution of types, which we can use to generate the same statistics as in the paper. The most illustrative is to compute the usage response in response to a linear tariff (as in Table IV). When doing that, we find that the simple model underestimates the price response by an average of 38.6%, and predicts usage levels that are 28.4% different, on average.

S.2. ECONOMETRIC DETAILS

S.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 ($t = 1, 2, \dots, 30$ over a billing cycle with $T = 30$ days) for our daily data. These discretizations leave v_t as the only continuous state variable. Because the subscriber does not know v_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, $E[V_{hkt}(C_{t-1})]$, and policy functions, $E[c_{hkt}^*(C_{t-1})]$. To perform the

numerical integration over the bounded support of v_t , $[0, \bar{v}]$, 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 ($2,000 \times 30$) are implicitly defined by threshold values for v_t . For example, consider a subscriber of type h on plan k , that has consumed C_{t-1} prior to period t . The threshold, $v_t(z)$, is defined as the value of v_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 average charges) of an additional unit of consumption

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

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,

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

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.

S.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 $\widehat{\mathbf{m}}_k^{\text{dat}}$, 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,

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

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.

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

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

where $i \in \{1, \dots, N_{kt}(C_{t-1})\}$ 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 $\widehat{E}[c_{kt}^*(C_{t-1})]$ 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 $8 \times 120,000 = 960,000$ moments in total.

S.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 9, we present the marginal distribution of all five parameters using only plan selection and applying uniform weights to all types that choose a particular plan, 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.

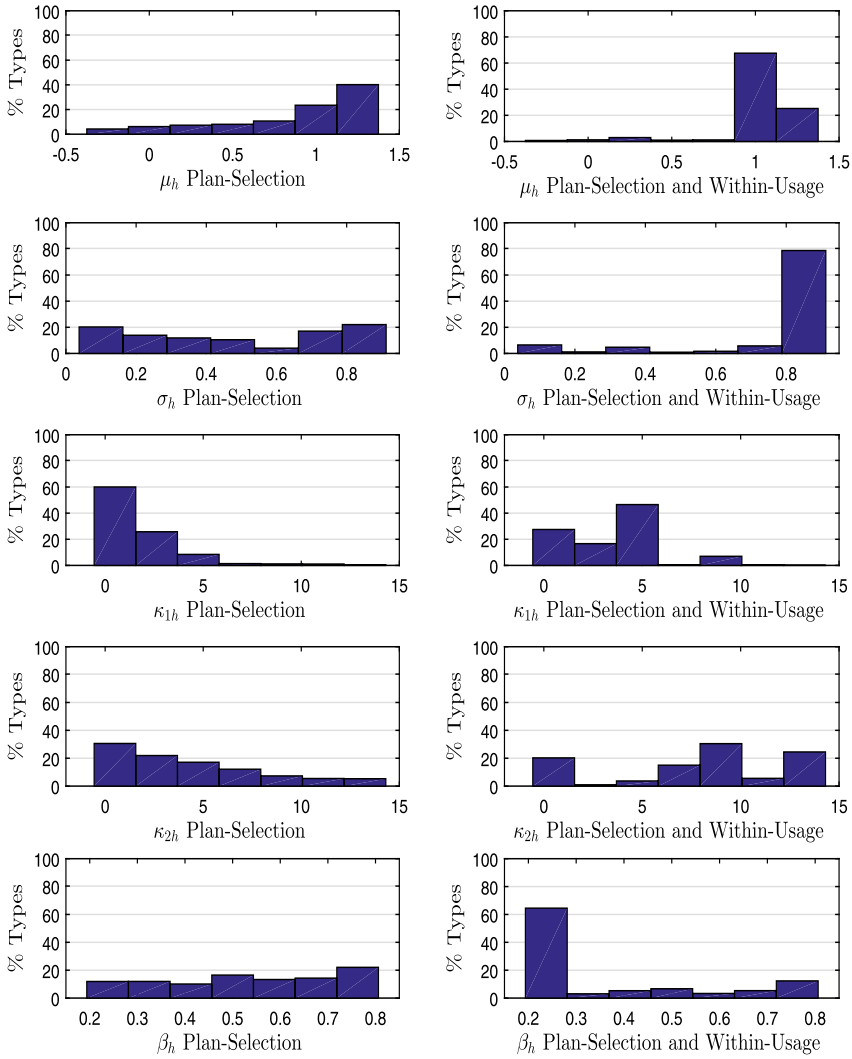


FIGURE 9.—Sources of identification: plan selection and usage. *Note:* This figure presents the marginal distribution of each parameter, when only information on optimal plan selection is used and uniform weights are applied, and when the weights are chosen using information on optimal plan selection and to match usage moments from the data.

S.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 10 and 11, we plot the model's predicted behavior for the four types with the greatest es-

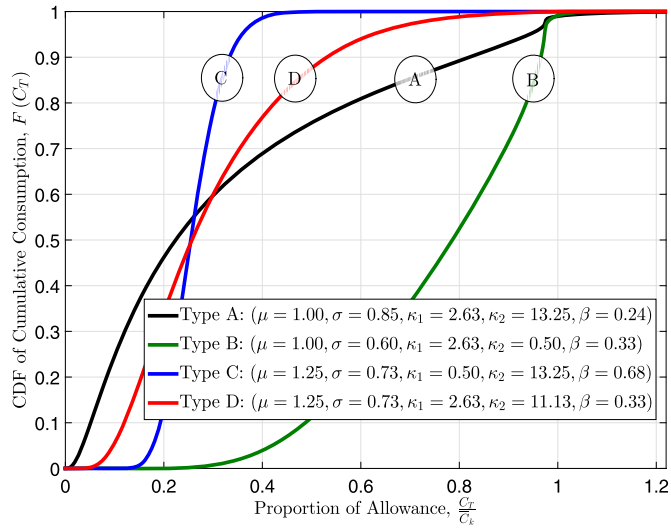


FIGURE 10.—Predicted behavior by type, CDF of C_T . *Note:* This figure presents the CDF of cumulative consumption on the last day of the billing cycle, C_T , for the types that received the greatest estimated weight on one of the usage-based plans in our data. The CDF is a function of C_T , but since this is for a single type, there is a one-to-one mapping to $\frac{c_T}{c_k}$.

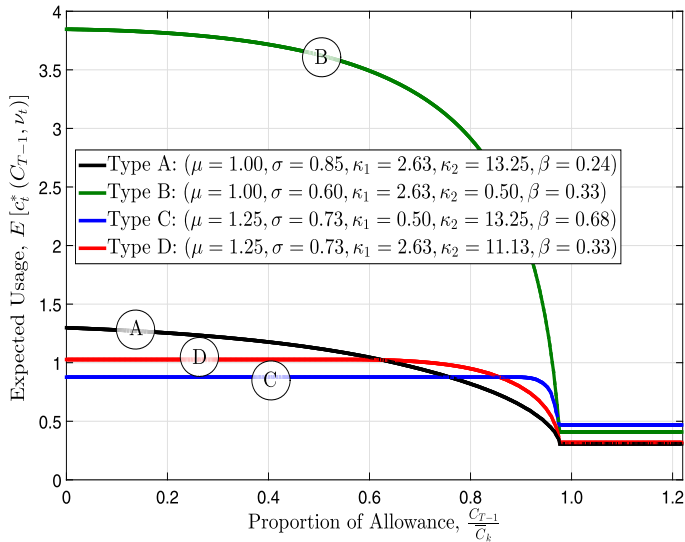


FIGURE 11.—Predicted behavior by type: expected usage $E[c_T^*(C_{T-1}, v_t)]$. *Note:* This figure presents expected usage on the last day of the billing cycle, $E[c_T^*(C_{T-1}, v_t)]$, for the types that received the greatest estimated weight on one of the usage-based plans in our data.

timated weights on a particular usage-based plan. Figure 10 plots the CDF of cumulative consumption on the final day of the billing cycle, C_T , for each of the types. Figure 11 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} . The functions in this figure highlight price sensitivity, through the change in expected usage as the fraction of the allowance used, $\frac{C_{T-1}}{C_k}$, increases. This price sensitivity also reveals information about variation in usage. A subscriber like type B in Figure 11 whose cumulative usage is well below the allowance, but whose expected usage changes with small movements in the fraction of the allowance used, has some chance of very high usage.

Figures 10 and 11 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 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 12 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 13 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} .

For each of the parameters, perturbations result in quite different behavioral responses. An increase in μ results in the entire distribution of cumulative consumption shifting to the right, that is, 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, that is, 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, that is, the frequency of very low and very high cumulative consumption states is reduced. Figure 13 shows that perturbations of each of the parameters has a differing effect on expected consumption conditional on

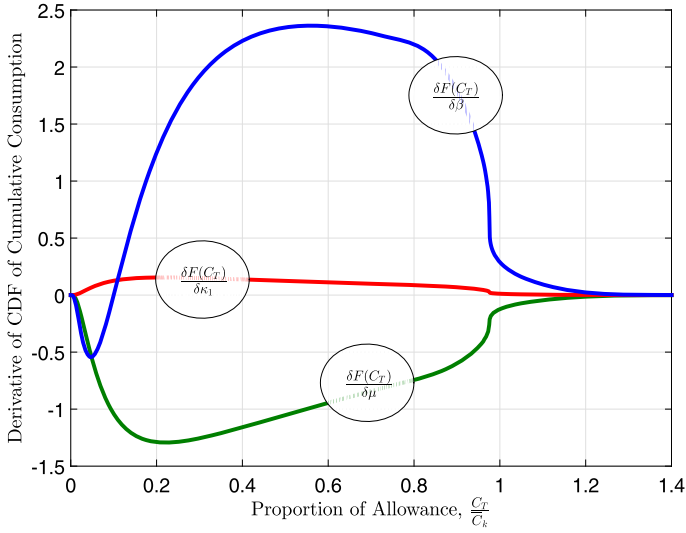


FIGURE 12.—Perturbation of parameters, CDF of C_T . *Note:* This figure presents the derivative of the CDF of cumulative usage on the last day of the billing cycle with respect to each of the type-specific parameters, μ , κ_1 , and β , for the most common type, ($\mu = 1.00$, $\sigma = 0.60$, $\kappa_1 = 2.63$, $\kappa_2 = 0.50$, $\beta = 0.33$). The derivative is a function of C_T , but since this is for a single type, there is a one-to-one mapping to $\frac{C_T}{C_k}$.

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.

S.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 $7^5 = 16,807$ types. This grid was chosen carefully through extensive experimentation to ensure that the type space we considered was “centered” over the types that place some value on broadband, and economic restrictions on the parameters. For example, β is naturally restricted between zero and 1 by economic theory, while there is a reasonable range of κ_1 that is bounded below by zero and captures an individual’s time cost.

The grid increments in the state space are arbitrary, and so are the number of moments. On one hand, too fine a grid and the number of observations available at the various state points becomes too small to compute a mean or variance (the smoothing across states, we describe above, helps). In addi-

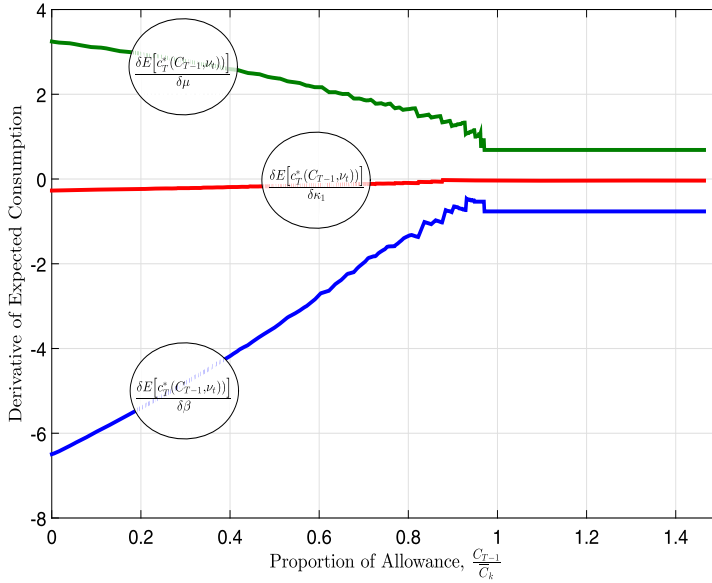


FIGURE 13.—Perturbation of parameters, expected usage. *Note:* This figure presents the derivative of expected consumption conditional on reaching a state on the last day of the billing cycle with respect to each of the type-specific parameters, μ , κ_1 , and β , for the most common type, ($\mu = 1.00$, $\sigma = 0.60$, $\kappa_1 = 2.63$, $\kappa_2 = 0.50$, $\beta = 0.33$).

tion, we cannot increase the density of the grid without severe multicollinearity problems (i.e., types behave too similarly at all states). On the other hand, too coarse a grid and the aggregation across states means information is lost to identify the parameters.

To show that our approach performs well even when the density of the grid of types is substantially reduced, we take the following approach. For each parameter, we consider the same range of support (minimum and maximum unchanged), but remove every other point. 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%.

The estimated distribution is slightly different, but we find the implications of the type distribution for overall demand are rather similar. To visualize this, Table X presents average daily usage under the same set of linear tariffs as Table IV. 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

TABLE X
 EXPECTED DAILY USAGE UNDER A LINEAR TARIFF, COARSE GRID^a

Price (\$)	Speed (Mb/s)				
	2 Mb/s	14.68 Mb/s	50 Mb/s	100 Mb/s	1,024 Mb/s
0.00	0.45 (0.005)	2.22 (0.010)	2.99 (0.015)	3.37 (0.018)	4.28 (0.034)
1.00	0.27 (0.002)	1.26 (0.001)	1.74 (0.005)	1.97 (0.006)	2.57 (0.009)
2.00	0.19 (0.001)	0.80 (0.002)	1.06 (0.003)	1.18 (0.003)	1.48 (0.004)
3.00	0.14 (0.001)	0.54 (0.001)	0.70 (0.001)	0.77 (0.001)	0.94 (0.001)
4.00	0.11 (0.001)	0.39 (0.001)	0.49 (0.001)	0.53 (0.001)	0.63 (0.001)
5.00	0.09 (0.001)	0.29 (0.001)	0.36 (0.001)	0.38 (0.001)	0.45 (0.001)

^aThis 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. Standard errors, in parentheses, are calculated using a block-resampling methodology as described in the text.

complex and higher-dimensional problems, where a less-dense grid of types is necessary.

Dept. of Economics, Northwestern University, 2001 Sheridan Rd., Evanston, IL 60208, U.S.A. and NBER; nevo@northwestern.edu,

Dept. of Economics, University of Georgia, Brooks Hall, Athens, GA 30602, U.S.A.; jturner@uga.edu,

and

Dept. of Economics, University of North Carolina–Chapel Hill, Gardner Hall, Chapel Hill, NC 27599, U.S.A.; jonwms@unc.edu.

Co-editor Liran Einav handled this manuscript.

Manuscript received September, 2013; final revision received October, 2015.