Web Appendix

accompanying

Consumer Dynamic Usage Allocation and Learning under Multi-part Tariffs

Arun Gopalakrishnan

Raghuram Iyengar

Robert J. Meyer

Appendix A: Propositions relating to optimal policy and threshold rules

In this Appendix, we first characterize the properties of the value function, dropping the subscripts for individual *i* and game *g* for notational convenience.

Proposition 1: The expected continuation value function is non-decreasing in $\underline{u}_t$ at given values of $1 \leq D_t < L$ and $0 \leq w_t < t$.

For $t = T$, $\overline{V_T}\left(\underline{u}'_T, D_T, w_T\right) = \int \max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), \underline{\mathrm{u}}'_{\mathrm{T}} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), 0\} f(u_T)\, du_T$

and $\overline{V_T}\left(\underline{u}''_T, D_T, w_T\right) = \int \max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), \underline{\mathrm{u}}''_{\mathrm{T}} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), 0\} f(u_T)\, du_T$

If $\underline{u}''_T \geq \underline{u}'_T$, $\max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), \underline{\mathrm{u}}''_{\mathrm{T}} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), 0\} \geq \max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), \underline{\mathrm{u}}'_{\mathrm{T}} - \mathrm{cI}(\mathrm{w_T} \geq \overline{\mathrm{W}}), 0\}\ \forall\, u_T$.

Hence $\overline{V_T}\left(\underline{u}''_T, D_T, w_T\right) \geq \overline{V_T}\left(\underline{u}'_T, D_T, w_T\right)$.

Assume $\bar{V}_{t+1}\left(\underline{u}''_{t+1}, D_{t+1}, w_{t+1}\right) \geq \bar{V}_{t+1}\left(\underline{u}'_{t+1}, D_{t+1}, w_{t+1}\right)$ whenever $\underline{u}''_{t+1} \geq \underline{u}'_{t+1}$.

Given $\overline{V_t}\left(\underline{\mathrm{u}}_{\mathrm{t}}, \mathrm{D_t}, \mathrm{w_t}\right) = \int \max\{\mathrm{u_t} - \mathrm{cI}(\mathrm{w_t} \geq \overline{\mathrm{W}}) + \beta \cdot \bar{V}_{t+1}(\mathrm{u_t}, L-1, w_t + 1), \underline{\mathrm{u}}_{\mathrm{t}} - \mathrm{cI}(\mathrm{w_t} \geq \overline{\mathrm{W}}) + \beta \cdot \bar{V}_{t+1}(\underline{\mathrm{u}}_{\mathrm{t}}, D_t - 1, w_t + 1), \beta \cdot \bar{V}_{t+1}(0,0,\mathrm{w_t})\} f(u_t)\, du_t$,

$\overline{V_t}\left(\underline{\mathrm{u}}_{\mathrm{t}}'', \mathrm{D_t}, \mathrm{w_t}\right) \geq \overline{V_t}\left(\underline{\mathrm{u}}_{\mathrm{t}}', \mathrm{D_t}, \mathrm{w_t}\right)$ for $\underline{u}''_t \geq \underline{u}'_t$, since $\underline{\mathrm{u}}_{\mathrm{t}} - \mathrm{cI}(\mathrm{w_t} \geq \overline{\mathrm{W}}) + \beta \cdot \bar{V}_{t+1}(\underline{\mathrm{u}}_{\mathrm{t}}, D_t - 1, w_t + 1)$ is a non-decreasing function of $\underline{\mathrm{u}}_{\mathrm{t}}$ (by assumption for $\bar{V}_{t+1}$ term and obvious for $\underline{\mathrm{u}}_{\mathrm{t}}$ linear term). By induction, the proposition holds true $\forall\, t$.

Proposition 2: The expected continuation value function is non-increasing in w_t at given values of $\underline{u}_t$ and $1 \leq D_t < L$.

For $t = T$, $\overline{V_T}\left(\underline{u}_T, D_T, w'_T\right) = \int \max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w}'_{\mathrm{T}} \geq \overline{\mathrm{W}}), \underline{u}_T - \mathrm{cI}(\mathrm{w}'_{\mathrm{T}} \geq \overline{\mathrm{W}}), 0\} f(u_T)\, du_T$

and $\overline{V_T}\left(\underline{u}_T, D_T, w''_T\right) = \int \max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w}''_{\mathrm{T}} \geq \overline{\mathrm{W}}), \underline{u}_T - \mathrm{cI}(\mathrm{w}''_{\mathrm{T}} \geq \overline{\mathrm{W}}), 0\} f(u_T)\, du_T$

If $w''_T \geq w'_T$, $\max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w}'_{\mathrm{T}} \geq \overline{\mathrm{W}}), \underline{u}_T - \mathrm{cI}(\mathrm{w}'_{\mathrm{T}} \geq \overline{\mathrm{W}}), 0\} \leq \max\{\mathrm{u_T} - \mathrm{cI}(\mathrm{w}''_{\mathrm{T}} \geq \overline{\mathrm{W}}), \underline{u}_T - \mathrm{cI}(\mathrm{w}''_{\mathrm{T}} \geq \overline{\mathrm{W}}), 0\}\ \forall\, u_T$, with equality holding when both w'_T and w''_T are strictly less or greater than $\overline{\mathrm{W}}$.

Hence $\overline{V_T}\left(\underline{u}_T, D_T, w''_T\right) \leq \overline{V_T}\left(\underline{u}_T, D_T, w'_T\right)$.

Assume $\bar{V}_{t+1}\left(\underline{\mathrm{u}}_{\mathrm{t+1}}, D_{t+1}, w''_{t+1}\right) \leq \bar{V}_{t+1}\left(\underline{\mathrm{u}}_{\mathrm{t+1}}, D_{t+1}, w'_{t+1}\right)$ whenever $w''_{t+1} \geq w'_{t+1}$.

Given $\overline{V_t}\left(\underline{\mathrm{u}}_{\mathrm{t}}, \mathrm{D_t}, \mathrm{w_t}\right) = \int \max\{\mathrm{u_t} - \mathrm{cI}(\mathrm{w_t} \geq \overline{\mathrm{W}}) + \beta \cdot \bar{V}_{t+1}(\mathrm{u_t}, L-1, w_t + 1), \underline{\mathrm{u}}_{\mathrm{t}} - \mathrm{cI}(\mathrm{w_t} \geq \overline{\mathrm{W}}) + \beta \cdot \bar{V}_{t+1}(\underline{\mathrm{u}}_{\mathrm{t}}, D_t - 1, w_t + 1), \beta \cdot \bar{V}_{t+1}(0,0,\mathrm{w_t})\} f(u_t)\, du_t$,

$\overline{V_t}\big(\underline{u}_t, D_t, w_t''\big) \le \overline{V_t}\big(\underline{u}_t, D_t, w_t'\big)$ for $w_t'' \ge w_t'$, since $-cI(w_t \ge \overline{W}) + \beta \cdot \bar{V}_{t+1}(u_{t+1}, D_{t+1}, w_t + 1)$ is a non-increasing function of w_t (by assumption for $\bar{V}_{t+1}$ term and obvious for $-cI(w_t \ge \overline{W})$ term). By induction, the proposition holds true $\forall\, t$.

Proposition 3: The expected continuation value function is non-decreasing in D_t at given values of $\underline{u}_t$ and $0 \le w_t < t$.

$1 \le D_t < L$ in our setting, since $D_t = 0$ implies no ongoing call, and $D_t \ge L$ is not possible since this implies that the call could exceed the maximum call duration of L.

For $t = T$, $\overline{V_T}\big(\underline{u}_T, D_T'', w_T\big) = \overline{V_T}\big(\underline{u}_T, D_T', w_T\big)$ for any admissible value of D_T'', D_T'. With only one time period left, it does not matter how many additional periods of talk time are left on an ongoing call as long as there is at least one period. It is therefore also true that $\overline{V_T}\big(\underline{u}_T, D_T'', w_T\big) \ge \overline{V_T}\big(\underline{u}_T, D_T', w_T\big)$ for $D_T'' \ge D_T'$.

Assume $\bar{V}_{t+1}\big(\underline{u}_{t+1}, D_{t+1}'', w_{t+1}\big) \ge \bar{V}_{t+1}\big(\underline{u}_{t+1}, D_{t+1}', w_{t+1}\big)$ whenever $D_{t+1}'' \ge D_{t+1}'$.

Given $\overline{V_t}\big(\underline{u}_t, D_t, w_t\big) = \int \max\{u_t - cI(w_t \ge \overline{W}) + \beta \cdot \bar{V}_{t+1}(u_t, L-1, w_t+1), \underline{u}_t - cI(w_t \ge \overline{W}) + \beta \cdot \bar{V}_{t+1}(\underline{u}_t, D_t - 1, w_t + 1), \beta \cdot \bar{V}_{t+1}(0,0,w_t)\} f(u_t)\, du_t$,

$\bar{V}_t\big(\underline{u}_t, D_t'', w_t\big) \ge \bar{V}_t\big(\underline{u}_t, D_t', w_t\big)$ since $\bar{V}_{t+1}(\underline{u}_t, D_t - 1, w_t + 1)$ is non-decreasing in D_t by assumption and is the only term that involves D_t. By induction, the proposition holds true $\forall\, t$.

A consequence of Proposition 1 and 2 is that $\bar{V}_{t+1}(0,0,w_t) - \bar{V}_{t+1}(\underline{u}_t, D_t - 1, w_t + 1)$, which features in both equations (2) and (3) is not generally greater than or less than zero, since the expected continuation value function $\bar{V}_{t+1}(\underline{u}_t, D_t - 1, w_t + 1)$ has both a weakly higher continuing call utility and greater used talk time.

Proposition 4: Switching to a call of equal or higher utility than an existing call is always better than continuing on the existing call.

Switching makes sense when $u_t + \beta \cdot \bar{V}_{t+1}(u_t, L-1, w_t+1) \ge \underline{u}_t + \beta \cdot \bar{V}_{t+1}\big(\underline{u}_t, D_t - 1, w_t + 1\big)$. If $u_t \ge \underline{u}_t$, and given $L \ge D_t$,

$\bar{V}_{t+1}(u_t, L-1, w_t+1) \ge \bar{V}_{t+1}\big(\underline{u}_t, D_t - 1, w_t + 1\big)$ by Proposition 1 and 3. Hence, switching weakly dominates continuation. Note that it can, in some circumstances, be optimal to switch to a call of *lower* utility than an existing call. Hence, Proposition 4 establishes a sufficient but not necessary condition for switching.

We now describe the threshold rules that apply based on the consumer's decision space at each time period. In equation (1), the value function at time *t* depends on which of the four decision scenarios manifests in that time period. The continuation value of accepting a new call

$\bar{V}_{ig,t+1}(u_{igt}, L-1, w_{igt}+1)$ reflects the extra second of talk time used at time *t* while having an ongoing call with utility u_{igt} and remaining duration $L-1$ (since L is the maximum duration of a call in our decision problem). Similarly, the continuation value of staying on an existing call $\bar{V}_{ig,t+1}(\underline{u}_{igt}, D_{igt}-1, w_{igt}+1)$ reflects the extra second of talk time used at time *t* and call progress. If the consumer rejects or terminates a call, or has no decision to make, the continuation value $\bar{V}_{ig,t+1}(0,0,w_{igt})$ reflects that talk time stays the same and no ongoing call will be available in the next period.

The optimal policy depends on the available decisions. If the only decision is to accept or reject a new call (no ongoing call), a threshold rule applies. As shown in equation (A1), the optimal decision rule ($\delta^*_{igt,ACC-REJ}$) is either 1 (accept) or -1 (reject). The intuition is straightforward: overage cost raises the bar for accepting calls, a higher continuation value for taking a call lowers the bar, and a higher continuation value for call rejection raises the bar. The difference in continuation values is critical to the optimal policy as this weighs the saving of another second of usage against the utility from continuing the call in the future. The effect of the tradeoff on the decision rule depends crucially on the amount of talk time used (w_{igt}).

$$\delta^*_{igt,ACC-REJ}(u_{igt}, w_{igt}) = 2 \cdot I\left(u_{igt} > c_{ig}\mathrm{I}\left(w_{igt} \geq \bar{\mathrm{W}}_{ig}\right) + \beta_{ig} \cdot \bar{V}_{ig,t+1}(0,0,w_{igt}) - \beta_{ig} \cdot \bar{V}_{ig,t+1}(u_{igt}, L-1, w_{igt}+1)\right) - 1 \quad \text{(A1)}$$

Similarly, another threshold rule applies when the only option is to continue ($\delta^*_{igt,CON-TER} = 0$) or terminate ($\delta^*_{igt,CON-TER} = -1$) an existing call as shown in equation (A2). The intuition is analogous to the new call acceptance decision, except that the utility and remaining duration are based on an existing call.

$$\delta^*_{igt,CON-TER}(\underline{u}_{igt}, D_{igt}, w_{igt}) = I\left(\underline{u}_{igt} > c_{ig}\mathrm{I}\left(w_{igt} \geq \bar{W}_{ig}\right) + \beta_{ig} \cdot \bar{V}_{ig,t+1}(0,0,w_{igt}) - \beta_{ig} \cdot \bar{V}_{ig,t+1}(\underline{u}_{igt}, D_{igt}-1, w_{igt}+1)\right) - 1 \quad \text{(A2)}$$

When both a new and existing call are available, the consumer's optimal decision involves two steps. The first step determines whether switching to the new call ($\delta^*_{igt,SWITCH} = 1$) or staying with the existing call ($\delta^*_{igt,SWITCH} = 0$) is more attractive (equation A3). The more attractive option is then weighed against the continuation value of declining all calls, resulting in $\delta^*_{igt,SWITCH-CON-TER}$ being a composite of the decision rules in equations (A1), (A2), and (A3).

$$\delta^*_{igt,SWITCH}(u_{igt}, \underline{u}_{igt}, D_{igt}, w_{igt}) = I\left(u_{igt} + \beta_{ig} \cdot \bar{V}_{ig,t+1}(u_{igt}, L-1, w_{igt}+1) \geq \underline{u}_{igt} + \beta_{ig} \cdot \bar{V}_{ig,t+1}(\underline{u}_{igt}, D_{igt}-1, w_{igt}+1)\right) \quad \text{(A3)}$$

$$\begin{aligned}\delta^*_{igt,SWITCH-CON-TER}(u_{igt}, \underline{u}_{igt}, D_{igt}, w_{igt}) &= \delta^*_{igt,CON-TER}(\underline{u}_{igt}, D_{igt}, w_{igt}) \cdot \left[1 - \delta^*_{igt,SWITCH}(u_{igt}, \underline{u}_{igt}, D_{igt}, w_{igt})\right] + \\ &\delta^*_{igt,ACC-REJ}(u_{igt}, w_{igt}) \cdot \delta^*_{igt,SWITCH}(u_{igt}, \underline{u}_{igt}, D_{igt}, w_{igt})\end{aligned} \quad \text{(A4)}$$

Appendix B: Examples of call choice data collected from participants and what in the data identify structural parameters

In Figures B1 and B2, we show illustrations of call arrivals (marked as circular points), usage (indicated by the solid lines – 0 denotes not being on a call at that time either due to lack of calls or deciding against taking/staying on a call), and when usage hits talk time allowance (dotted vertical line) from our first study (in the 40-second talk time condition). The subject in Figure B1, for example, rejects several high-value calls in the first two games and experiences overage for the first time in game three. The subject then allocates usage to all high-value calls but also takes lower-valued calls in the remaining three games (which may or may not be optimal). In contrast, another subject from the same experimental condition (Figure B2) always allocates usage to high-value calls and tends to be conservative with using talk time allowance for lower-valued calls after game one. In game six, this subject turns down 13 consecutive lower-valued calls and starts using talk time close to the 50-second mark.

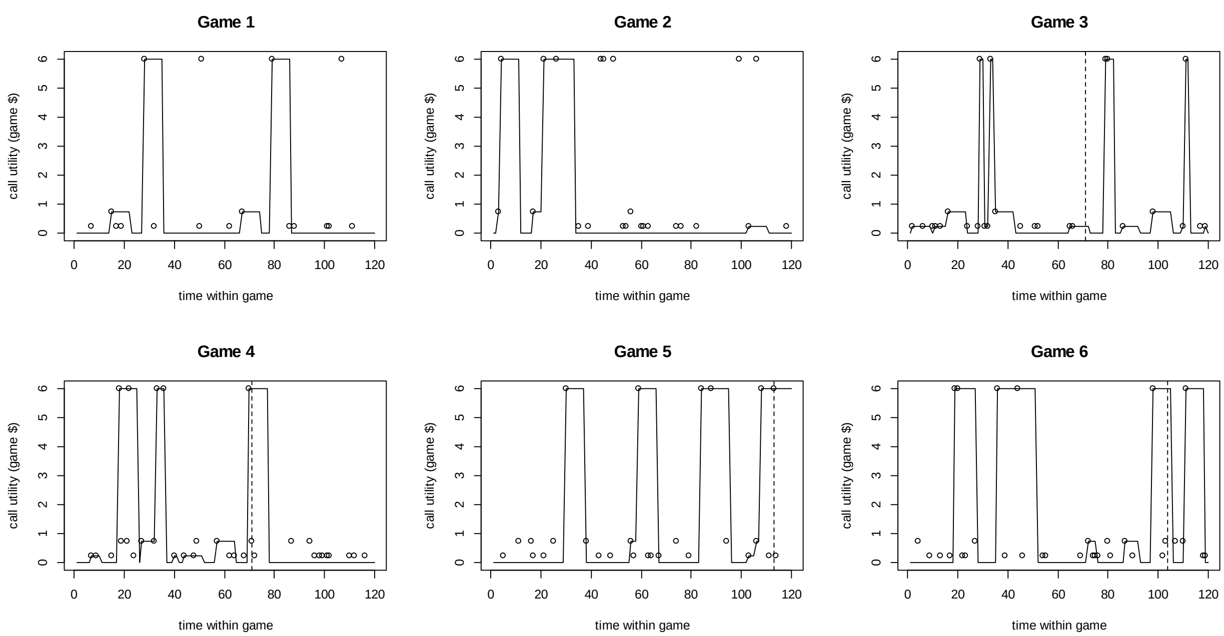


Figure B1: Illustrations of call arrival and usage patterns for a subject with large errors under a 40-second allowance plan for all six games. The dotted vertical line is when free call allowance was used up (if at all).

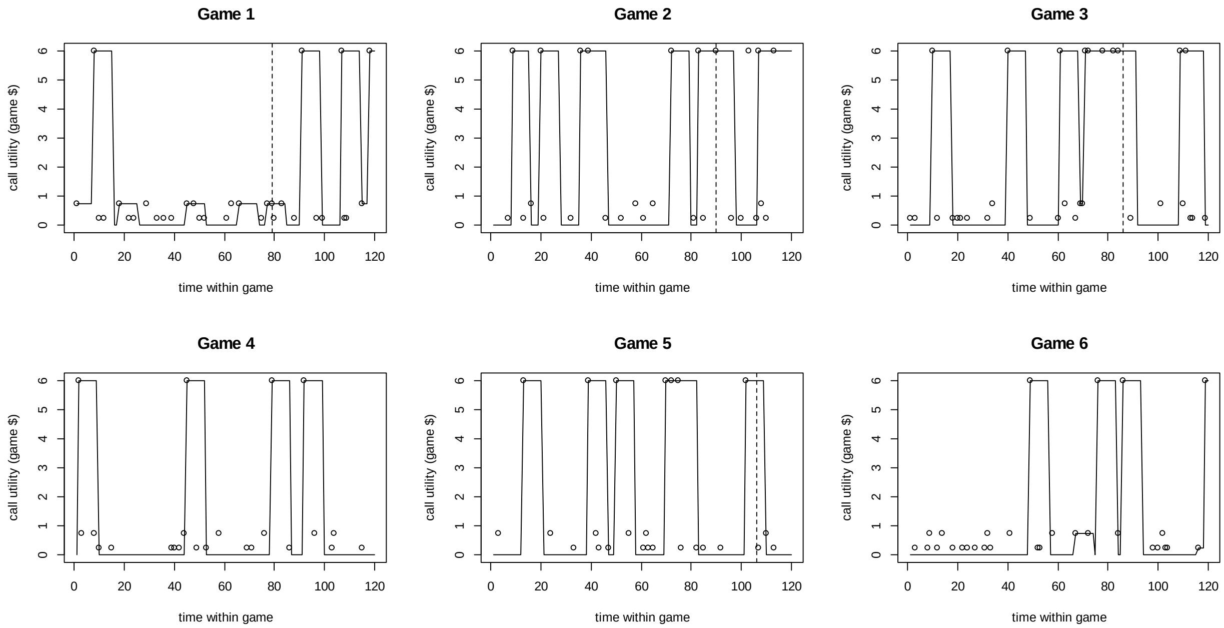


Figure B2: Illustrations of call arrival and usage patterns for a subject who learns to allocate usage to mostly high-valued calls under a 40-second allowance plan for all six games. The dotted vertical line is when free call allowance was used up (if at all).

Though plan and game characteristics have been chosen to allow for empirical recovery of structural parameters, estimator properties also depend on the "signal" contained in each game about an individual's decision process due to the stochastic nature of call arrivals. Imagine that a person received *only* high-value calls and accepted all of them. This would be consistent with many models of behavior (including a myopic consumer). However, the timing of acceptance or rejection of low- and medium-value calls (e.g., Figures B1 and B2) is informative of how forward looking a person may be. Specifically, if a person did take these lower-valued calls early on, the inflection points along consumed talk time (as illustrated in Figure 3) at which the person stops taking such calls helps identify the discount factor. Rejecting high-value calls is informative of a noisy policy (with large idiosyncratic shocks) as is exhibiting deviations from the optimal policy for any given discount factor (moderate idiosyncratic shocks).

Appendix C: Full set of Screenshots

In this appendix, we reproduce the exact instructions that subjects were exposed to, as they undertook the simulation.

Screen 1:

Welcome to the Call Study Simulator

On the next screen you will be introduced to a cell-phone call study in which you will make a series of decisions. Bonus money can be earned based on your performance.

To get started, enter your MTurk worker ID in the field below and press the Start button below. Upon completing the study, which will take 15 to 20 minutes, you will get a code to enter in the MTurk window from which you launched this study, to get paid. No payment will be made if the study is not completed.

Subject id:

Start

Screen 2:

Introduction to the calling game

This is a study about phone-call management. You will be playing a game where the objective is to maximize the amount of pleasure you get from calls while minimizing the cost of those calls. It works like this. You will play a game where you have to make a series of decisions about whether to accept or reject a series of incoming calls in a 120-second window. You will play this game 6 times.

At the start of the game, you will be randomly assigned to a payment plan which provides a certain amount of "free" calling time for a fixed up-front charge. If you go over this allowance you are charged by the second for all the calls you take. Unused free seconds do not roll over to subsequent games and the payment plan may change from game to game. Check the plan details which will appear at the top of the screen before playing each game.

There are three types of calls – low-value ($0.25/sec), medium-value ($0.75/sec) and high-value ($6.00/sec). At any given moment, you may receive one of these call types or no call. If you do receive a call, you will be asked if you want to accept or reject the call.

Each call lasts for a maximum of 8 seconds. If you accept a call, you can hang up anytime, or switch to another arriving call.

The probability of receiving a call at any point is 20%. If a call is received, 50% of the time it will be low-valued, 25% of the time it will be medium-valued, and 25% of the time, it will be high-valued.

All amounts shown during the task are in Game dollars, and your accumulated net value after the 6 games will be converted into real US currency using an exchange rate of $0.10 for every 100 Game dollars. The bonus range would be roughly in the range of US$0.40 to US$1, and is contingent on Game dollars accumulated. If your net value is negative, no bonus will be paid.

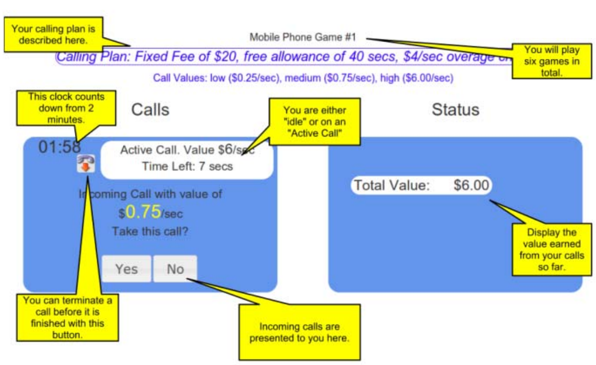
The image on the next page shows the main features of the game screen.

After the next screen, you will first play a 30 second practice game which does not count towards your net value. You will then start the first of the six games.

Press to see game screen

Screen 3:

Press to begin practice game



Screen 4:

Screen 5:

The following payment plan will apply for your call usage in game 1:

A fixed fee of $80,

an allowance of 40 free seconds, and

a $4.00 per second overage charge.

You will be notified if the payment plan changes.
Please check the top of the game screen to verify the payment plan for each game.

Click here to begin next game

Screen 6:

Appendix D: MCMC Algorithm for 2-segment latent class model

Markov Chain Monte Carlo (MCMC) procedures are used for obtaining posterior distributions of the following parameters: β_{gs} – the discount factor for segment *s* in game g, λ_{gs} – the scale parameter for segment *s* in game g, α – the probability of belonging to segment 1, Z_i – the latent class membership for each subject *i*. We transform the segment and game specific parameters as $\begin{bmatrix}\text{logit}(\beta_{gs})\\ \log(\lambda_{gs})\end{bmatrix} = \mu_{gs}$.

The joint distribution of all model parameters is given by:

$$p(\{\mu_{g1}, \mu_{g2}\ \forall g\}, \alpha, \{Z_i \forall i\}|Y)$$
$$\propto p(\alpha)\prod_{g=1}^{6} p(\mu_{g1})p(\mu_{g2})\prod_{i=1}^{N} p(Z_i|\alpha)\left(\prod_{g=1}^{6} p(Y_{ig}|\mu_{g1})\right)^{I(Z_i=1)}\left(\prod_{g=1}^{6} p(Y_{ig}|\mu_{g2})\right)^{I(Z_i=2)}$$

The following steps will implement a hybrid Gibbs/Metropolis sampler.

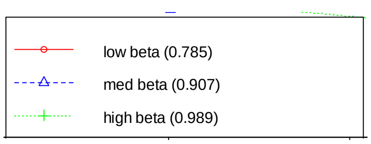
1. Draw from full condition distribution of $\mu_{gs}|\{Z_i\forall i\}, Y\ \ \forall g, s$:
   $p(\mu_{gs})\prod_{i=1}^{N}\left(p(Y_{ig}|\mu_{gs})\right)^{I(Z_i=s)}$ using a Metropolis step since the distribution has no closed form. $p(Y_{ig}|\mu_{gs})$ requires a backward sweep to compute expected value functions that are used to form conditional choice probabilities. $p(\mu_{gs})$ is uniformly distributed in the region where $\text{logit}(\beta_{gs}) \in [-10,10]$ and $\log(\lambda_{gs}) \in [-10,10]$ and has zero support outside this region. Due to the logit and log transformations, the range of discount factors covered is $[0.00005, 0.99995]$ and the range of error scale parameters is $[0.00005, 22{,}026.5]$. Values outside of this range would be extremely unlikely to materialize in this setting.
2. α conditional on $\{Z_i\forall i\}$ is deterministic and is computed as $\frac{\sum_{i=1}^{N} I(Z_i=1)}{N}$.
3. Draw $Z_i|\{\mu_{g1}, \mu_{g2}\ \forall g\}, \alpha, Y \sim Bernoulli\left(\frac{\alpha\left(\prod_{g=1}^{6} p(Y_{ig}|\mu_{g1})\right)}{\alpha\left(\prod_{g=1}^{6} p(Y_{ig}|\mu_{g1})\right)+(1-\alpha)\left(\prod_{g=1}^{6} p(Y_{ig}|\mu_{g2})\right)}\right)$.

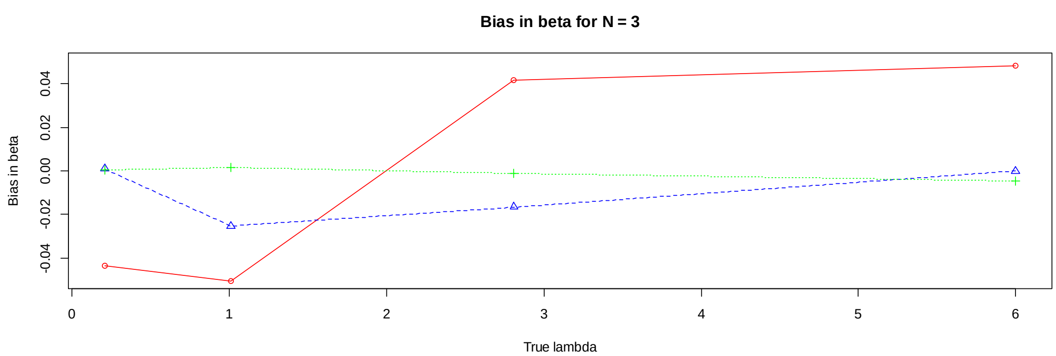
To generalize to S segments, steps 2 and 3 simply need to be modified to account for the additional segments. The most computationally intensive process (within step 1) is to calculate the *ex ante* value function at each point in state space and time for use in the conditional choice probabilities. We parallelize this computation for efficiency using Amazon Web Services (AWS) elastic cloud computing resources. We implemented this using the R library “parallel” which allows dynamic creation and allocation of tasks to worker processes spawned as new threads. Computationally intensive tasks were written in the C programming language and linked to the R interface as a shared library. The C and R source code for implementing this algorithm are available upon request from the authors. 100,000 iterations of the algorithm were run for each data set.

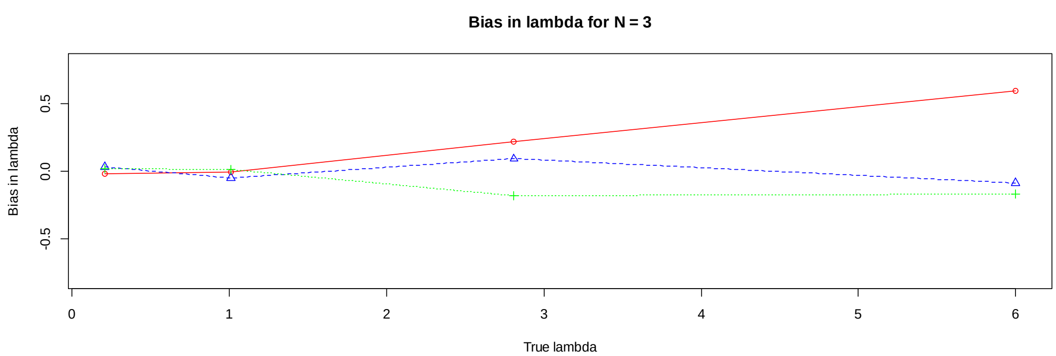
Appendix E: Simulation and Parameter Recovery for Model

We test parameter recovery for varying amounts of data for a grid of true $\beta \in \{0.785, 0.907, 0.989\}$ and $\lambda \in \{0.21, 1.01, 2.81, 6.00\}$ under the 40-second plan allowance. This set of parameter values covers the typical range of estimates that we find in our studies. For each true parameter set, we simulate 30 datasets of three, six, and twelve subjects (N) each by generating random call arrivals and applying the decision rule given by the behavioral model for the true parameters for each call decision. For the purposes of this simulation, each dataset is composed of homogenous subjects with one game each. The finite-sample bias is computed by estimating the posterior mean of the parameters for each of the data sets and computing the average deviation from the true parameters. We observe that for N = 3, recovery of parameters is more robust for higher true β. N = 6 provides a much tighter parameter recovery by comparison and for N = 12, recovery is fairly tight around the true parameters. The implication for latent-class analysis is that any class should have at least 6 subjects to allow for robust parameter recovery.

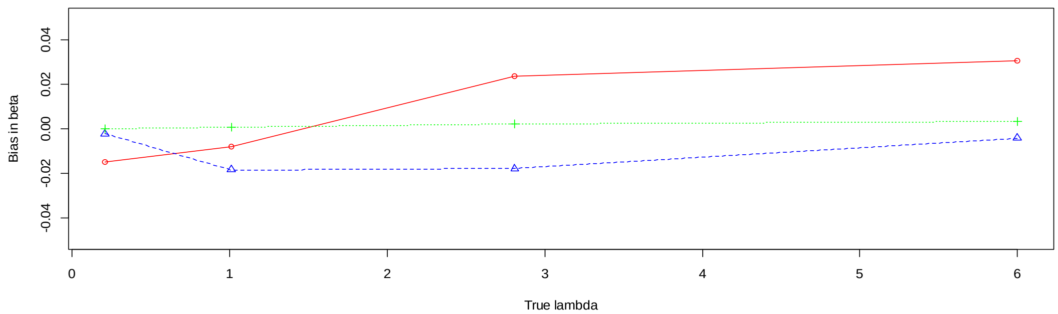
The legend below applies for all plots.



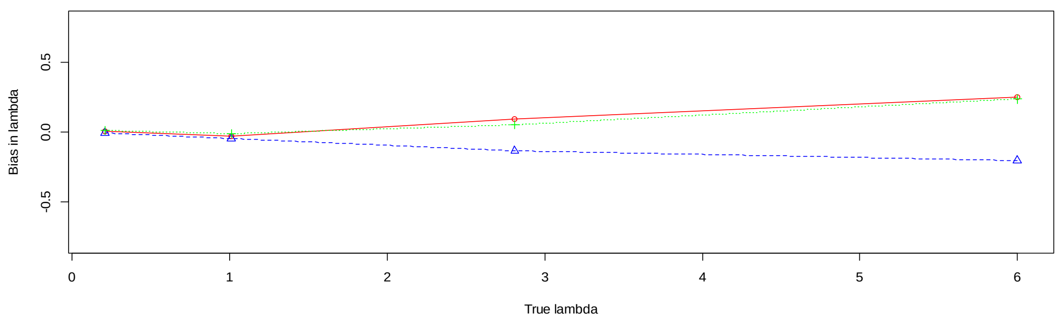




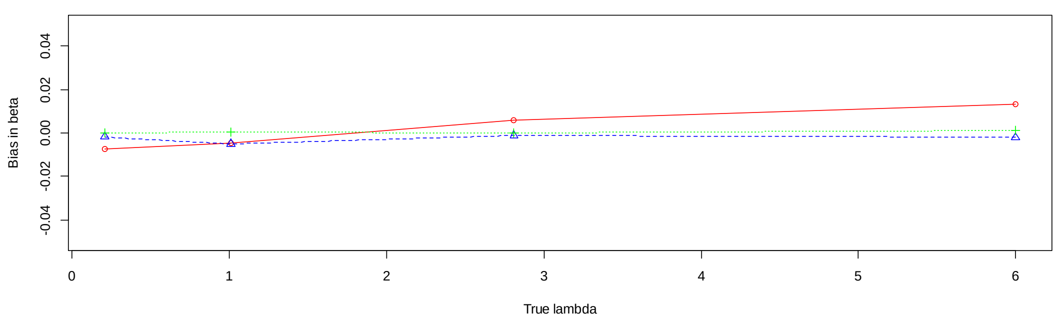
Bias in beta for N = 6



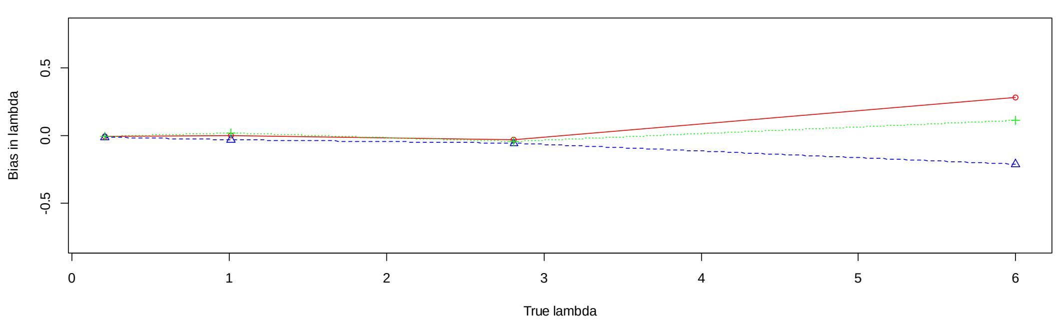
Bias in lambda for N = 6



Bias in beta for N = 12



Bias in lambda for N = 12



Appendix F: Model Comparison Results for Study 1

In Study 1, we estimate a set of models for the 20-second and 40-second plan allowance conditions to characterize fit . We find that a 2-segment latent class model allowing for learning in both β (discount factor) and λ (scale parameter representing deviations from optimal behavior) fits well compared to a homogenous model or 2-segment models in which learning is turned off for either parameter. While increasing the number of latent classes beyond two segments does improve fit, the largest improvement occurs when comparing a homogenous model to one that assumes two segments. The model fit results for the other studies are similar. Since our goal is to characterize the effects of three-part tariffs on overall task performance, we use 2 latent classes for ease of comparison across studies, and these capture much of the distinct differences between two highly heterogeneous sub-groups.

	1 segment	2 segments	3 segments	4 segments
Total Parameters	12	25	37	51
Log Marginal Likelihood (LML)	-6,264	-4,967	-4,568	-4,283
DIC	12,528	9,930	9,142	8,574

Table F1: Study 1 20-second plan allowance condition (N = 44) – comparing fit across different numbers of latent classes

	No Learning on β, No Learning on λ	Learning on β, No Learning on λ	No Learning on β, Learning on λ	Learning on β, Learning on λ
Total Parameters	5	15	15	25
Log Marginal Likelihood (LML)	-5,739	-5,532	-5,006	-4,967
DIC	11,477	11,050	10,003	9,930

Table F2: Study 1 20-second plan allowance condition (N = 44) – comparing fit across nested 2-segment models.

	1 segment	2 segments	3 segments	4 segments
Total Parameters	12	25	37	51
Log Marginal Likelihood (LML)	-9,592	-8,608	-8,223	-7,943
DIC	19,185	17,200	16,395	15,896

Table F3: Study 1 40-second plan allowance condition (N = 63) – comparing fit across different numbers of latent classes

	No Learning on β, No Learning on λ	Learning on β, No Learning on λ	No Learning on β, Learning on λ	Learning on β, Learning on λ
Total Parameters	5	15	15	25
Log Marginal Likelihood (LML)	-9,428	-9,185	-8,811	-8,608
DIC	18,855	18,366	17,626	17,200

Table F4: Study 1 40-second plan allowance condition (N = 63) – comparing fit across nested 2-segment models.

Appendix G: Details of Study 4 procedure

In Study 4, we administered the Holt and Laury (2002) lottery task prior to our main cellphone simulation task as shown in the screenshot below.

Part I - Lottery Task

For each of the pairs of choice below, choose the option you prefer (either Option A or B). There is no right or wrong answer.

For example, 5%: $1.00, 95%: $0.50 would mean that in a hypothetical lottery, you would have a 5% chance of winning $1.00 and a 95% chance of winning $0.50.

Only pick one option per choice. Once you are done with all 10 choices, click the submit button.

Choice 1:	10%: $2.00, 90% $1.60	10%: $3.85, 90%: $0.10
Choice 2:	20%: $2.00, 80% $1.60	20%: $3.85, 80%: $0.10
Choice 3:	30%: $2.00, 70% $1.60	30%: $3.85, 70%: $0.10
Choice 4:	40%: $2.00, 60% $1.60	40%: $3.85, 60%: $0.10
Choice 5:	50%: $2.00, 50% $1.60	50%: $3.85, 50%: $0.10
Choice 6:	60%: $2.00, 40% $1.60	60%: $3.85, 40%: $0.10
Choice 7:	70%: $2.00, 30% $1.60	70%: $3.85, 30%: $0.10
Choice 8:	80%: $2.00, 20% $1.60	80%: $3.85, 20%: $0.10
Choice 9:	90%: $2.00, 10% $1.60	90%: $3.85, 10%: $0.10
Choice 10:	100%: $2.00, 0% $1.60	100%: $3.85, 0%: $0.10

Submit

For this study, we recruited a new panel of 48 participants from Amazon Mechanical Turk, using the same recruitment criteria as Study 2. From the original pool, 5 were screened out based on a pre-defined criterion of task comprehension (accepting 33% or fewer of all high-value calls across the six games). Of the remaining 43 subjects, two participants' lottery task responses displayed inconsistent preferences. Specifically, one participant accepted option A for choice 10 which is inconsistent with any level of risk aversion since a certain $3.85 dominates a certain $2.00. Another participant switched from option A to B, and then back again, which Holt and Laury (2002) also designate as violating the basic premise of the task. These two participants were removed from the dataset to yield a final pool of 41 participants.

A risk neutral participant would pick option A for choices 1 through 4 and switch to option B from choice 5 onwards. We code risk attitude as the number of option A choices less 4 (such that 0 corresponds to risk neutral, > 0 corresponds to risk averse, and < 0 corresponds to risk seeking). On average, subjects are risk averse in line with Holt and Laury (2002).