Online Appendix for

Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock

Michael D. Grubb and Matthew Osborne

A Data Details

A.1 Price Data

Table 7: Popular Plan Price Menu

	Plan 0 ($14.99)				Plan 1 ($34.99)				Plan 2 ($44.99)				Plan 3 ($54.99)			
Date	Q	p	OP	Net	Q	p	OP	Net	Q	p	OP	Net	Q	p	OP	Net
8/02 - 10/02	-	-	-	-	280	40	free	not	653	40	free	not	875	35	free	not
10/02 - 12/02	**0**	**11**	**free**	**free**	280	40	free	not	653	40	free	not	875	35	free	not
12/02 - 1/03	0	11	free	free	**350**	40	free	not	653	40	free	not	875	35	free	not
1/03 - 2/03	0	11	free	free	**280**	40	free	not	653	40	free	not	875	35	free	not
2/03 - 3/03	0	11	free	free	***380***	40	free	not	653	40	free	not	875	35	free	not
3/03 - 9/03	0	11	free	free	**288**	**45**	free	not	**660**	40	free	not	**890**	**40**	free	not
9/03 - 1/04	0	11	**not**	free	**388**	45	free	not	660	40	free	not	890	40	free	not
1/04 - 4/04	0	11	not	free	388	45	free	not	660	40	free	**free**	890	40	free	not
4/04 - 5/04	0	11	not	free	388	45	free	not	**1060**	40	free	free	890	40	free	not
5/04 - 7/04	0	11	not	free	**288**	45	free	not	**760**	40	free	free	890	40	free	not

Entries describe the calling allowance (Q), the overage rate (p), whether off-peak calling is free or not (OP), and whether in-network calling is free or not (Net). Bold entries reflect price changes that apply to new plan subscribers. The Bold italics entry reflects the one price change which also applied to existing plan subscribers. Some terms remained constant: Plan 0 always offered $Q = 0$, $p = 11$, and free in-network. Plans 1-3 always offered free off-peak.

Prices of the four popular plans are described for all dates in Table 7.The variation in the number of minutes offered and other plan terms provides us with a rich source of variation in the data. However, it also created difficulties in constructing price data, as we were not provided with the schedule of promotions. For each plan and each date, we infer the total number of included free minutes by observing the number of minutes used prior to an overage in the call-level data. (This calculation is complicated by the fact that some plans offered free in-network calls, and our call-level data does not identify whether an incoming call was in-network.) We were unable to reliably infer this pricing information at all dates for plans with small customer shares, which is why we group unpopular plans (national plans, free-long-distance plans, and expensive local plans) with the outside option in our structural model. These plans comprised only 11% of all bills.

A subscriber's calling plan is identified by a rate-plan-code recorded on each subscriber's bill, and the date the subscriber chose the plan. The later determines promotional features such as free in-network calling or free bonus minutes applicable to the plan. The date a left censored subscriber chose their plan is unobserved, and hence so are the plan terms.

A.2 Sample Selection

Bills were available from the dates of February 2002 to June of 2005 and contain billing information for 2334 subscribers. We exclude roughly 500 individuals who enroll before February 2002 and are therefore left censored. We do so both because their initial plan choices are unobserved and because terms of their February 2002 plan choices are unobserved. For our structural estimation and reduced form analysis in Section 3.2.2, we also restrict attention to the period August 2002 to July 2004. We first remove bills prior to June 2002 because adjustment to a new billing system caused some problems in the data during this period. Second, we exclude data after October 2004 because after this time the university stopped offering service to new customers and began encouraging existing subscribers to transfer management of their accounts directly to the cellular-phone provider. Third we restrict attention further the period August 2002 to July 2004 because this is the period in which we can reliably infer popular plan prices from billing data. Combined with other sample restrictions described in the main text the restrictions mean that our final data set contains 1366 subscribers and 16,283 month-subscriber observations. Reduced form analysis in Section 3.2.1 that is focused on usage choice rather than plan choice (and hence is less reliant on observing the complete price menu at every date) uses bills from a slightly wider sample period.

B Model Details

B.1 Derivation of optimal calling threshold

Define $q(p,\theta_{it}^k) \equiv \arg\max_q \left(V\left(q,\theta_{it}^k\right) - pq\right)$ to be a consumer's demand for category-k calls given a constant marginal price p. (This is the quantity of category-k calls valued above p.) A consumer's inverse demand for category-k calls is $V_q(q_{it}^k,\theta_{it}^k) = \left(\theta_{it}^k/q_{it}^k - 1\right)/\beta$ and thus:

$$q(p,\theta_{it}^k) = \theta_{it}^k/\left(1+\beta p\right) = \theta_{it}^k \hat{q}(p). \tag{12}$$

Conditional on tariff choice j with free off-peak calling, consumer i chooses her period t peak threshold v_{it}^{pk} to maximize her expected utility conditional on her period t information $\Im_{it}$:

$$v_{itj}^{pk} = \arg\max_{v^*} E\left[V\left(q(v^*,\theta_{it}^{pk}),\theta_{it}^{pk}\right) - P_j\left(q(v^*,\theta_{it}^{pk})\right) \mid \Im_{it}\right].$$

Let $\tilde{F}_{it}$ be the cumulative distribution of θ_{it}^{pk} as perceived by consumer i at time t. The first-order condition for the consumer's problem is

$$\int_{\underline{\theta}}^{\bar{\theta}} V_q\left(q(v^*,\theta_{it}^{pk}),\theta_{it}^{pk}\right)\frac{d}{dv^*}q(v^*,\theta_{it}^{pk})d\tilde{F}_{it}(\theta_{it}^{pk}) = \int_{\theta_j^*(v^*)}^{\bar{\theta}} p_j\frac{d}{dv^*}q(v^*,\theta_{it}^{pk})d\tilde{F}_{it}(\theta_{it}^{pk}), \tag{13}$$

where $\theta_j^*(v^*)$ is the peak type which consumes exactly Q_j units: $q(v^*,\theta_j^*(v^*)) = Q_j$. Equation (13) is similar to Borenstein's (2009) first-order condition. Unlike Borenstein (2009), we assume $V_q(q(v^*,\theta_{it}^k),\theta_{it}^k)$ is equal to v^* by definition,[49] so this condition reduces to:

$$v^* = p_j\frac{\int_{\theta_j^*(v^*)}^{\bar{\theta}}\frac{d}{dv^*}q(v^*,\theta_{it}^{pk})d\tilde{F}_{it}(\theta_{it}^{pk})}{\int_{\underline{\theta}}^{\bar{\theta}}\frac{d}{dv^*}q(v^*,\theta_{it}^{pk})d\tilde{F}_{it}(\theta_{it}^{pk})}.$$

With multiplicative separability (equation (12)), $\theta_j^*(v^*) = Q_j/\hat{q}(v^*)$ and $\frac{d}{dv^*}q(v^*,\theta_{it}^{pk}) = \theta_{it}^{pk}\frac{d}{dv^*}\hat{q}(v^*)$, so we can factor out and cancel $\frac{d}{dv^*}\hat{q}(v^*)$. This yields equation (2). It is apparent by inspection that this has a unique solution.

B.2 Optimal calling threshold under bill-shock regulation

For simplicity we temporarily drop the peak superscript as well as customer, date, and plan subscripts. Given bill-shock regulation, a consumer will begin the billing cycle consuming all peak calls above some threshold v^* but raise this calling threshold to the overage rate p upon receiving a bill-shock warning that the allowance Q has been reached.

Let θ^* be the peak type which consumes exactly Q units given initial calling threshold v^*: $q(v^*,\theta^*) = Q$. If $\theta < \theta^*$ then the customer has a total of θ calling opportunities, makes $\theta\hat{q}(v^*) < Q$ calls with value $V(q(v^*,\theta),\theta)$ and never receives a bill-shock warning. If $\theta > \theta^*$, then prior to receiving a bill-shock warning the customer has θ^* calling opportunities and makes $\theta^*\hat{q}(v^*) = Q$ calls with total value $V(Q,\theta^*)$. After receiving a bill-shock warning, the customer has $\theta^a = \theta - \theta^*$ additional calling opportunities and makes $\theta^a\hat{q}(p)$ calls with total value $V(q(p,\theta^a),\theta^a)$.

The consumer chooses v^* to maximize

$$U = -M + \int_0^{\theta^*} V(q(v^*,\theta),\theta) f(\theta)\, d\theta + \int_{\theta^*}^{\infty}\left(V(Q,\theta^*) + V(q(p,\theta^a),\theta^a) - pq(p,\theta^a)\right) f(\theta)\, d\theta,$$

where $\theta^* = Q/\hat{q}(v^*)$ and $\theta^a = \theta - \theta^*$. After cancelling terms (recognizing that $q(v^*,\theta^*) = Q$, $q(p,0) = 0$, and $V(0,\theta) = 0$) and substituting $v^* = V_q(q(v^*,\theta),\theta)$ and $d\theta^a/dv^* = -d\theta^*/dv^*$, the

[49]Subsituting $q_{it}^k = q\left(v^*,\theta_{it}^k\right) = \theta_{it}^k/(1+\beta v^*)$ into $V_q(q_{it}^k,\theta_{it}^k)$ yields v^*.

first derivative is

$$\frac{dU}{dv^*} = \int_0^{\theta^*} v^* \frac{d}{dv^*} q(v^*, \theta) f(\theta) d\theta + \frac{d\theta^*}{dv^*} \int_{\theta^*}^{\infty} (V_\theta(Q, \theta^*) - V_\theta(q(p, \theta^a), \theta^a)) f(\theta) d\theta.$$

The multiplicative demand assumption implies $\frac{d}{dv^*} q(v^*, \theta) = \theta \frac{d}{dv^*} \hat{q}(v^*)$ and $d\theta^*/dv^* = -(\theta^*/\hat{q}(v^*)) \frac{d}{dv^*} \hat{q}(v^*)$. Therefore the $\frac{d}{dv^*} \hat{q}(v^*)$ terms cancel and the first-order condition is

$$v^* \int_0^{\theta^*} \theta f(\theta) d\theta = \frac{\theta^*}{\hat{q}(v^*)} \int_{\theta^*}^{\infty} (V_\theta(Q, \theta^*) - V_\theta(q(p, \theta^a), \theta^a)) f(\theta) d\theta.$$

Substituting $V_\theta(q, \theta) = (\ln(q/\theta) - 1)/\beta$ yields

$$v^* \int_0^{\theta^*} \theta f(\theta) d\theta = \frac{1}{\beta} \frac{\theta^*}{\hat{q}(v^*)} (\ln(\hat{q}(v^*)) - \ln(\hat{q}(p))) (1 - F(\theta^*)), \tag{14}$$

which characterizes v^* given $\theta^* = Q/\hat{q}(v^*)$ and $\hat{q}(v^*) = 1/(1 + \beta v^*)$.

B.3 Bayesian Updating

At the end of the first billing period $t = 1$, consumer i learns $z_{i1} = (1 - \varphi) \tilde{\theta}_{i1}^{pk}$, which she believes has distribution

$$N\left(\mu_i^{pk}, \frac{1 - \varphi}{1 + \varphi} (\tilde{\sigma}_\varepsilon^{pk})^2\right).$$

Then, in later periods $t > 1$, consumer i learns $z_{it} = \tilde{\theta}_{it}^{pk} - \varphi \tilde{\theta}_{i,t-1}^{pk}$, which she believes has distribution $N(\mu_i^{pk}, (\tilde{\sigma}_\varepsilon^{pk})^2)$. Define $\bar{z}_{it} = \frac{1}{t} \sum_{\tau=1}^{t} z_{i\tau}$. Then by Bayes rule (DeGroot 1970), updated time $t + 1$ beliefs about μ_i^{pk} are $\mu_i^{pk} | \Im_{i,t+1} \sim N(\tilde{\mu}_{i,t+1}^{pk}, \tilde{\sigma}_{t+1}^2)$ where

$$\tilde{\mu}_{i,t+1}^{pk} = \frac{\tilde{\mu}_{i1}^{pk} \tilde{\sigma}_1^{-2} + \left(\frac{2\varphi}{1-\varphi} z_{i1} + t\bar{z}_{it}\right) (\tilde{\sigma}_\varepsilon^{pk})^{-2}}{\tilde{\sigma}_1^{-2} + \left(\frac{2\varphi}{1-\varphi} + t\right) (\tilde{\sigma}_\varepsilon^{pk})^{-2}}, \tag{15}$$

and

$$\tilde{\sigma}_{t+1}^2 = \left(\tilde{\sigma}_1^{-2} + \left(\frac{2\varphi}{1 - \varphi} + t\right) (\tilde{\sigma}_\varepsilon^{pk})^{-2}\right)^{-1}.$$

Note that equation (15) can be re-written to show that a consumer's updated beliefs are a weighted average of her prior and her signals, where the weight placed on her prior is proportional to $(\tilde{\sigma}_\varepsilon^{pk}/\tilde{\sigma}_1)^2$:

$$\tilde{\mu}_{i,t+1}^{pk} = \frac{(\tilde{\sigma}_\varepsilon^{pk}/\tilde{\sigma}_1)^2 \tilde{\mu}_{i1}^{pk} + \left(\frac{2\varphi}{1-\varphi} z_{i1} + t\bar{z}_{it}\right)}{(\tilde{\sigma}_\varepsilon^{pk}/\tilde{\sigma}_1)^2 + \left(\frac{2\varphi}{1-\varphi} + t\right)}. \tag{16}$$

B.4 Biases

Holding the distribution of $\{\tilde{\mu}_{i1}^{pk}, \mu_i^{pk}, \mu_i^{op}\}$ fixed, description of bias must be relative to a chosen rational-expectations benchmark. Our chosen benchmark is $\tilde{\mu}_{i1}^{pk} = E[\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}]$ and $\tilde{\sigma}_1^2 = Var(\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk})$. An alternative would be $\tilde{\mu}_{i1}^{pk} = E[\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op}]$ and $\tilde{\sigma}_1^2 = Var(\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op})$, as we assume consumers know their off-peak types. Applying the standard formula for a conditional distribution from a joint-normal distribution, the conditional distribution $\mu_i^{pk} \mid \{\tilde{\mu}_{i1}^{pk}, \mu_i^{op}\}$ has mean

$$E\left[\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op}\right] = \mu_0^{pk} + \psi^{pk}\left(\tilde{\mu}_{i1}^{pk} - \tilde{\mu}_0^{pk}\right) + \frac{\sigma_{\mu^{pk}}}{\sigma_{\mu^{op}}}\rho_\mu\left(\mu_i^{op} - \mu_0^{op} - \psi^{op}\left(\tilde{\mu}_{i1}^{pk} - \tilde{\mu}_0^{pk}\right)\right)$$

and variance $Var(\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op}) = \left(1-\rho_\mu^2\right)\sigma_{\mu^{pk}}^2$. Defining $\delta'_\mu = \tilde{\sigma}_1 / \sqrt{Var(\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op})}$, from the variance we have $\delta'_\mu = \delta_\mu / \sqrt{1-\rho_\mu^2}$. Rearranging terms and adding and subtracting $\tilde{\mu}_0^{pk}$ to the mean shows that

$$\tilde{\mu}_{i1}^{pk} - E\left[\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op}\right] = \left(\tilde{\mu}_0^{pk} - \mu_0^{pk}\right) + \left(1 - \psi^{pk} + \psi^{op}\rho_\mu\frac{\sigma_{\mu^{pk}}}{\sigma_{\mu^{op}}}\right)\left(\tilde{\mu}_{i1}^{pk} - \tilde{\mu}_0^{pk}\right) - \rho_\mu\frac{\sigma_{\mu^{pk}}}{\sigma_{\mu^{op}}}\left(\mu_i^{op} - \mu_0^{op}\right).$$

Defining $b'_2 = 1 - \psi^{pk} + \psi^{op}\rho_\mu\left(\sigma_{\mu^{pk}}/\sigma_{\mu^{op}}\right)$ and $b_3 = -\rho_\mu\left(\sigma_{\mu^{pk}}/\sigma_{\mu^{op}}\right)$, this is equivalent to:

$$\tilde{\mu}_{i1}^{pk} - E\left[\mu_i^{pk} \mid \tilde{\mu}_{i1}^{pk}, \mu_i^{op}\right] = b_1 + b'_2(\tilde{\mu}_{i1}^{pk} - \tilde{\mu}_0^{pk}) + b_3\left(\mu_i^{op} - \mu_0^{op}\right).$$

Thus this alternative rational-expectations benchmark implies an adjustment to the definitions of δ_μ and b_2 as well as a new dimension of bias measured by b_3.

Our estimate of b_3 is -0.447 (0.005), which implies that if consumers really knew their off-peak type, they would be underreacting to that information when forming beliefs about their peak type. However, we do not interpret this as an additional dimension of bias, but rather the fact that consumers do not actually know their off-peak types.[50] As a result, our favored rational-expectations benchmark is that given in the main text, which does not condition on μ_i^{op}. Note that while the choice of rational-expectations benchmark directly effects measures of bias, it does not effect the economic consequences (such as the value of bill-shock alerts) of an estimated distribution of $\{\tilde{\mu}_{i1}^{pk}, \mu_i^{pk}, \mu_i^{op}\}$. Thus it does not effect any of our welfare results except for our de-biasing counterfactual simulations.

[50] We need to make an assumption about beliefs about off-peak type when we estimate the model, because those beliefs affect consumer decisions to choose plan 0 with costly off-peak usage after September 2003. Because variation in off-peak plan prices is insufficiently rich to identify off-peak beliefs, for simplicity we assume that consumers do know their off-peak type during estimation. This assumption is innocuous for our results because off-peak usage is almost always free in estimation and our fixed-price counterfactuals and is always exogenously free in our endogenous-price counterfactuals.

C Complete Model

The complete model differs from the illustrative version presented in the main text by accounting for free in-network calling. Calling plans distinguish between in-network and out-of-network calls as well as between peak and off-peak calls. Thus the usage vector has four dimensions rather than two,

$$\mathbf{q}_{it} = (q_{it}^{pk,out}, q_{it}^{pk,in}, q_{it}^{op,out}, q_{it}^{op,in}),$$

where the superscript notation is: (1) "pk" for peak calls, (2) "op" for off-peak calls, (3) "out" for out-of-network calls, and (4) "in" for in-network calls. Popular pricing plans are the same function of total billable minutes as before, but billable minutes now depend on an indicator for whether the plan charges for in-network calls (NET_j):

$$q_{itj}^{billable} = q_{it}^{pk,out} + NET_j q_{it}^{pk,in} + OP_j q_{it}^{op,out} + NET_j OP_j q_{it}^{op,in}.$$

The peak and off-peak taste shock vector $\boldsymbol{\theta}_{it}$ follows the same process as in the illustrative model. In addition, we incorporate the taste shock $\mathbf{r}_{it} = (r_{it}^{pk}, r_{it}^{op}) \in [0,1]^2$ which captures the share of peak and off-peak demand that is for in-network calling rather than out-of-network calling. Together, $\boldsymbol{\theta}_{it}$ and $\mathbf{r}_{it}$ determine category specific taste shocks $\mathbf{x}_{it}$:

$$\mathbf{x}_{it} = \begin{bmatrix} x_{it}^{pk,out} \\ x_{it}^{pk,in} \\ x_{it}^{op,out} \\ x_{it}^{op,in} \end{bmatrix} = \begin{bmatrix} (1 - r_{it}^{pk})\theta_{it}^{pk} \\ r_{it}^{pk}\theta_{it}^{pk} \\ (1 - r_{it}^{op})\theta_{it}^{op} \\ r_{it}^{op}\theta_{it}^{op} \end{bmatrix}.$$

We continue to assume that calls from different categories are neither substitutes nor complements. Consumer i's utility from choosing plan j and consuming $\mathbf{q}_{it}$ in period t is thus

$$\begin{aligned} u_{itj} &= \sum_k V(q_{it}^k, x_{it}^k) - P_j(\mathbf{q}_{it}) + \frac{1}{\alpha}\eta_{itj}, \\ k &\in \{\text{pk-in, pk-out, op-in, op-out}\}. \end{aligned}$$

Moreover, we continue to assume that consumers choose separate calling thresholds for each type of call,

$$\mathbf{v}_{itj}^* = (v_{itj}^{pk,out}, v_{itj}^{pk,in}, v_{itj}^{op,out}, v_{itj}^{op,in}),$$

and, at the end of the month, realized usage in category k is $q_{it}^k = x_{it}^k \hat{q}(v^k)$. Implicitly this assumes that consumers can distinguish in-network and out-of-network numbers when choosing to make a call. In reality, consumers likely can't always do so but they likely can for parties they call in high volume. Finally, customer i's perceived expected utility from choosing plan j at date t is

$$U_{itj} = E\left[\sum_{k\in\{\text{pk-in,pk-out,op-in,op-out}\}} V\left(q^k(v^k_{itj}, x^k_{it}), x^k_{it}\right) - P_j\left(\mathbf{q}(\mathbf{v}^*_{itj}, \mathbf{x}_{it})\right) \mid \Im_{it}\right] + \frac{1}{\alpha}\eta_{itj}. \quad (17)$$

In general, the first-order conditions for threshold choice are analogous to the base model:

$$v^k_{it} = p_j \Pr\left(q^{total}_{itj} > Q_j\right) \frac{E\left[x^k_{it} \mid q^{total}_{itj} > Q_j\right]}{E\left[x^k_{it}\right]}.$$

Given the structure of the taste shocks, in-network and out-of-network thresholds only differ when in-network calls are free. There are four classes of tariff to consider. First, plan 0 prior to fall 2003 when both in-network and off-peak were free: $\mathbf{v}_{it} = (.11, 0, 0, 0)$. Second, plan 0 in fall 2003 or later when only in-network was free: $\mathbf{v}_{it} = (.11, 0, .11, 0)$. Third, three-part tariffs with free in-network calling, such as plan 2 in January 2004: $\mathbf{v}_{it} = (v^{pk,out}_{it}, 0, 0, 0)$ and

$$v^{pk,out}_{it} = p_j \Pr\left(x^{pk,out}_{it} \geq Q_j/\hat{q}(v^{pk,out}_{it})\right) \frac{E\left[x^{pk,out}_{it} \mid x^{pk,out}_{it} \geq Q_j/\hat{q}(v^{pk,out}_{it}); \Im_{it}\right]}{E\left[x^{pk,out}_{it} \mid \Im_{it}\right]}. \quad (18)$$

Fourth, standard three-part tariffs without free in-network calling: $\mathbf{v}_{it} = (v^{pk}_{it}, v^{pk}_{it}, 0, 0)$ and

$$v^{pk}_{it} = p_j \Pr\left(\theta^{pk}_{it} \geq Q_j/\hat{q}(v^{pk}_{it})\right) \frac{E\left[\theta^{pk}_{it} \mid \theta^{pk}_{it} \geq Q_j/\hat{q}(v^{pk}_{it}); \Im_{it}\right]}{E\left[\theta^{pk}_{it} \mid \Im_{it}\right]}.^{51} \quad (19)$$

As in the main text (Section 5), we break out calling demand for weekday outgoing calls to landlines immediately before and after 9pm to help identify the price coefficient. Now, however, such calls are a subset of out-of-network calling. Thus the shock $\mathbf{r}^{9pm}_{it} = (r^{9pk}_{it}, r^{9op}_{it}) \in [0,1]^2$ captures the share of peak and off-peak out-of-network calling demand that is within one hour of 9pm on a weekday and is for an outgoing call to a landline:

$$\mathbf{x}^{9pm}_{it} = \begin{bmatrix} x^{9pk}_{it} \\ x^{9op}_{it} \end{bmatrix} = \begin{bmatrix} r^{9pk}_{it} x^{pk,out}_{it} \\ r^{9op}_{it} x^{op,out}_{it} \end{bmatrix}.$$

[51] The expectation in equation (19) is taken with respect to the censored-normal θ^{pk}_{it}, and has a closed-form solution (up to the normal CDF, which has a standard built-in function in most programming languages). However, the expectation in equation (18) is taken with respect to the product of θ^{pk}_{it} and r^{pk}_{it} and does not have a closed-form solution. In the latter case we approximate the first-order condition using Gaussian quadrature.

Our identifying assumption that consumer i's expected outgoing calling demand to landlines on weekdays is the same between 8:00pm and 9:00pm as it is between 9:00pm and 10:00pm is unchanged, but this now corresponds to a revised restriction:

$$E\left[r_{it}^{9pk}\right]E\left[1-r_{it}^{pk}\right]E\left[\theta_{it}^{pk}\right]=E\left[r_{it}^{9op}\right]E\left[1-r_{it}^{op}\right]E\left[\theta_{it}^{op}\right], \tag{20}$$

in place of equation (7). Equation (20) implicitly defines α_i^{9op} as function of α_i^{9pk} and other parameters.

We model all calling share shocks r_{it}^k for $k \in \{pk, op, 9pk, 9op\}$ in the same manner as r_{it}^k for $k \in \{9pk, 9op\}$ in the illustrative model. We assume the distribution is a censored normal,

$$\begin{aligned}\tilde{r}_{it}^k &= \alpha_i^k + e_{it}^{r,k} \\ r_{it}^k &= \begin{cases} 0 & \text{if} \quad \tilde{r}_{it}^k \leq 0 \\ \tilde{r}_{it}^k & \text{if} \quad 0 < \tilde{r}_{it}^k < 1 \\ 1 & \text{if} \quad \tilde{r}_{it}^k \geq 1 \end{cases},\end{aligned}$$

where α_i^k is unobserved heterogeneity and $e_{it}^{r,k}$ is a mean-zero shock normally distributed with variance $(\sigma_e^k)^2$ independent across i, t, and k. For $k \in \{pk, op, 9pk\}$ we assume that α_i^k are normally distributed in the population (independently across i and k) with mean μ_α^k and variance $(\sigma_\alpha^k)^2$.

Beliefs about $\boldsymbol{\mu}_i$ and $\boldsymbol{\theta}_{it}$ are the same as in the illustrative model. We assume that there is no learning about the share of demand that is for in-network calling. Consumers know α_i^k and the distribution of $e_{it}^{r,k}$ for $k \in \{pk, op\}$ up to the fact that they underestimate the share of their calling opportunities that are in-network by a factor $\delta_r \in [0,1]$. ($\delta_r = 1$ corresponds to no bias.) Specifically, consumers believe that in-network calling shares have the distribution of $\delta_r \mathbf{r}_{it}$. We incorporate this additional bias to help explain consumers choice of plan 1 over plan 0, as plan 0 (with free in-network) dominates plan 1 (with costly in-network) at the median share of in-network calling. (Figure 6, which shows plan 1 undominated, corresponds to $\delta_r = 0$.)

C.1 Identification of Complete Model

Identifying beliefs involves one complication relative to the illustrative model: plan choice depends on beliefs about in-network calling shares as well as peak usage. A consumer's plan choice depends on her expected in-network peak-calling share $\delta_r E[r_{it}^{pk} \mid \alpha_i]$. Thus initial plan-choice shares depend on the population distribution of $\delta_r E[r_{it}^{pk} \mid \alpha_i]$. First consider a restricted model in which consumers are unbiased about in-network calling shares ($\delta_r = 1$). Then the population distribution of $\delta_r E[r_{it}^{pk} \mid \alpha_i]$ is identified without knowing beliefs using data prior to fall 2003. During this period, all plans offered free nights-and-weekends so that $r_{it}^{op} = q_{it}^{op,in}/q_{it}^{op}$. Moreover, only plan 0 offered free in-network calling. Thus for plans 1-3, peak calling-thresholds are the same for in-network and out-of-network calling and $r_{it}^{pk} = q_{it}^{pk,in}/q_{it}^{pk}$. For plan 0, peak calling-thresholds are 0 cents in-network and 11 cents out-of-network and hence

$$r_{it}^{pk} = \frac{q_{it}^{pk,in}}{q_{it}^{pk,in} + q_{it}^{pk,out}/\hat{q}\,(0.11)} = \frac{q_{it}^{pk,in}}{q_{it}^{pk,in} + (1+0.11\beta)\,q_{it}^{pk,out}},$$

where $1/\hat{q}\,(0.11) = (1+0.11\beta)$. Observing r_{it}^{k} for $k \in \{pk, op\}$ (a censoring of $\tilde{r}_{it}^{k} = \alpha_i^k + e_{it}^{r,k}$) identifies $E\left[\alpha_i^k\right]$, $Var(\alpha_i^k)$, and $Var(e_{it}^{r,k})$.[52]

When $\delta_r = 1$, the model is forced to rely on logit errors to explain plan 1's substantial share in fall 2002 and has trouble fitting the data, as plan 1 is dominated by plan 0 at the median in-network share. Hence we allow consumers to underestimate their in-network calling share. To separately identify δ_r from overconfidence (δ_μ) and volatility bias (δ_ε) it is important to use both pre and post fall-2003 plan-choice-shares. Reducing δ_r or reducing δ_μ and δ_ε both make plans 1-3 more favorable relative to plan 0. However, only δ_r has a differentially larger effect post fall 2003 when plan 0 stopped offering free nights-and-weekends. Thus the larger the drop in share of plan 0 between fall 2002 and fall 2003, the more fall 2002 plan 1 choices should be explained by low δ_r rather than overconfidence and volatility bias.

D Formulation of the Likelihood Function

In this section we outline some details relating to the construction of the likelihood that were omitted from Section 6. We begin by fully specifying $f_{\boldsymbol{q}}(\boldsymbol{q}_{it}|\boldsymbol{u}_i, \boldsymbol{q}_{i1}, ..., \boldsymbol{q}_{it-1}, \Theta)$. Recall that $\boldsymbol{q}_{it} = \{q_{it}^{pk,in}, q_{it}^{op,in}, q_{it}^{pk,out}, q_{it}^{op,out}, q_{it}^{9pk}, q_{it}^{9op}\}$ is a function of the random variables θ_{it}^{pk}, θ_{it}^{op}, and r_{it}^{k}, for $k \in \{pk, op, 9pk, 9op\}$. To compute $f_{\boldsymbol{q}}(\boldsymbol{q}_{it}|\boldsymbol{u}_i, \boldsymbol{q}_{i1}, ..., \boldsymbol{q}_{it-1}, \Theta)$, we first compute the likelihood of θ_{it}^{pk}, θ_{it}^{op}, and r_{it}^{k}, and then do the change of variables from these variables to $\boldsymbol{q}_{it}$.

Our first step is to back out $\boldsymbol{\theta}_{it,s}$ conditional on the threshold vector $\mathbf{v}_{it,s}^{*}$. This step is complicated somewhat by an important data limitation. We always observe total peak and off-peak calling $(q_{it}^{pk}, q_{it}^{op})$. However call logs only directly identify outgoing calls as in-network or out-of-network.

[52] A potential complication is that q_{it}^{pk-in} is only observed precisely for plan 0 subscribers and q_{it}^{op-in} is only observed precisely for fall 2003 and later subscribers to plan 0. For other plans we only observe bounds and a noisy estimate of q_{it}^{pk-in} because we can only distinguish in-network and out-of-network for outgoing calls. This measurement error problem is solvable because it only applies to a subset of the data. This comment also applies to r_{it}^{9op} and r_{it}^{9pk} which are now computed as: $r_{it}^{9op} = q_{it}^{9op}/q_{it}^{op,out}$ and $r_{it}^{9pk} = q_{it}^{9pk}/q_{it}^{pk,out}$.

This provides lower and upper bounds on in-network calling, $\underline{q}_{it}^{k,in}$ and $\bar{q}_{it}^{k,in}$ for $k \in \{pk, op\}$.[53] Fortunately, the network status of plan 0 peak calls (and off-peak calls for plan 0 that did not include free off-peak) can be inferred from whether they were charged 11 cents or 0 cents per minute. Thus, precisely when in-network calls are differentially priced, we can infer $q_{it}^{pk,in}$ and $q_{it}^{op,in}$ exactly.

For $k \in \{pk, op\}$, $\theta_{it,s}^{k}$ is calculated by equation (21) if category k calls are not priced differentially by network status or by equation (22) if category k calls are priced differentially by network status:

$$\theta_{it,s}^{k} = q_{it}^{k}/\hat{q}(v_{it,s}^{k}), \tag{21}$$

$$\theta_{it,s}^{k} = q_{it}^{k,in}/\hat{q}(v_{it,s}^{k,in}) + q_{it}^{k,out}/\hat{q}(v_{it,s}^{k,out}). \tag{22}$$

Latent $\tilde{\theta}_{it,s}^{k}$ equals $\theta_{it,s}^{k}$ when it is positive. Otherwise $\tilde{\theta}_{it,s}^{k}$ cannot be calculated due to censoring and we draw $\tilde{\theta}_{it,s}^{k}$ in period t from a conditional normal density.

When $\tilde{\boldsymbol{\theta}}_{it}$ is observed for both peak and off peak, its density is bivariate normal. This is because the structural error $\boldsymbol{\varepsilon}_{it}$ has a bivariate normal distribution conditional on simulated draws of $\boldsymbol{\mu}_{i,s} = (\mu_{i,s}^{pk}, \mu_{i,s}^{op})$. Therefore the AR1 process implies for $t > 1$

$$\tilde{\boldsymbol{\theta}}_{it,s} - \varphi\tilde{\boldsymbol{\theta}}_{it-1,s} \sim N\left(\boldsymbol{\mu}_{i,s}, \boldsymbol{\Sigma}_{\varepsilon,i}\right),$$

and for $t = 1$

$$\tilde{\boldsymbol{\theta}}_{i1,s} \sim N\left(\boldsymbol{\mu}_{i,s}/(1-\varphi), \boldsymbol{\Sigma}_{\varepsilon,i}/(1-\varphi^2)\right).$$

We denote the density of $\tilde{\boldsymbol{\theta}}_{it}$ described above as $f^{\theta}(\tilde{\boldsymbol{\theta}}_{it,s}|\tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta)$ for the bivariate normal density when $\tilde{\boldsymbol{\theta}}_{it,s}$ is not censored.

If either total-peak or total-off-peak usage is zero for some i and t, then the likelihood is adjusted to account for the censoring. When off-peak usage is zero but peak usage is positive, the likelihood is

$$\Pr(\tilde{\theta}_{it,s}^{op} < 0|\tilde{\theta}_{it,s}^{pk}, \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) f^{\theta,1}(\tilde{\theta}_{it,s}^{pk}|\tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta),$$

where $f^{\theta,1}$ denotes the univariate density of peak or off-peak usage conditional on past usage and the draws. When only peak usage is zero, we make a similar adjustment to the likelihood,

$$\Pr(\tilde{\theta}_{it,s}^{pk} < 0|\tilde{\theta}_{it,s}^{op}, \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) f^{\theta,1}(\tilde{\theta}_{it,s}^{op}|\tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta).$$

[53]The lower bound on total in-network usage is simply the total outgoing in-network minutes we observe. The upper bound is outgoing in-network minutes plus all incoming minutes.

When both peak and off-peak usage are zero, the likelihood is simply

$$Pr(\tilde{\theta}^{pk}_{it,s} < 0, \tilde{\theta}^{op}_{it,s} < 0 | \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta).$$

As described in the main text, when q^{pk}_{it} or q^{op}_{it} are zero we must draw the corresponding negative value of $\tilde{\theta}^{k}_{it}$. We follow the procedure developed by Lee (1999) for estimating dynamic tobit models. If peak usage is censored and off-peak usage is positive, we draw $\tilde{\theta}^{pk}_{it,s}$ from $f^{\theta,2}(\tilde{\theta}^{pk}_{it,s} | \tilde{\theta}^{pk}_{it,s} < 0, \tilde{\theta}^{op}_{it,s}, \tilde{\theta}_{it-1,s}, \boldsymbol{\mu}_{i,s}, t)$, where the density $f^{\theta,2}$ represents the truncated univariate normal density of θ^{pk}_{it} conditional on θ^{op}_{it}, prior period usage and simulated draws. The case when only off-peak usage is censored is symmetric. When both peak and off peak usage are zero, we draw both θ^{pk}_{it} and θ^{op}_{it} from a truncated bivariate normal distribution. This can be accomplished easily through importance sampling. (For an overview see Train (2009) pages 210-211.)

We can put this together to create a likelihood of $(\theta^{pk}_{it}, \theta^{op}_{it})$. Let the indicator $I^{k}_{it} = 1$ if total calls in category k are zero, for $k = \{pk, op\}$, and denote the vector of $\mathbf{v}^*_{it,s}$'s and $\tilde{\boldsymbol{\theta}}_{it,s}$'s for an individual i as $\mathbf{v}^*_{i,s}$ and $\tilde{\boldsymbol{\theta}}_{i,s}$ respectively. Then for some time period t and individual i, the likelihood of $(\theta^{pk}_{it}, \theta^{op}_{it})$ is

$$\begin{aligned} l^{\theta}(\tilde{\boldsymbol{\theta}}_{it,s}, I^{pk}_{it}, I^{op}_{it} | \boldsymbol{\mu}_{i,s}, \mathbf{v}^*_{i,s}, \Theta) \quad = \quad & \Big(f^{\theta}(\tilde{\theta}_{it,s} | \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) \Big)^{(1-I^{pk}_{it})(1-I^{op}_{it})} \\ & \Big(Pr(\tilde{\theta}^{pk}_{it,s} < 0 | \tilde{\theta}^{op}_{it,s}, \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) f^{\theta,1}(\tilde{\theta}^{op}_{it,s} | \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) \Big)^{I^{pk}_{it}(1-I^{op}_{it})} \\ & \Big(Pr(\tilde{\theta}^{op}_{it,s} < 0 | \tilde{\theta}^{pk}_{it,s}, \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) f^{\theta,1}(\tilde{\theta}^{pk}_{it,s} | \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) \Big)^{I^{op}_{it}(1-I^{pk}_{it})} \\ & \Big(Pr(\tilde{\theta}^{pk}_{it,s} < 0, \tilde{\theta}^{op}_{it,s} < 0 | \tilde{\boldsymbol{\theta}}_{i,t-1,s}, \boldsymbol{\mu}_{i,s}, t, \Theta) \Big)^{I^{pk}_{it} I^{op}_{it}} . \end{aligned} \tag{23}$$

Next we turn to the likelihood of in-network and out-of-network usage and 8:00 pm to 10:00 pm shares. We discuss the computation of this part in two pieces, first focusing on the in-network shares and then moving on to 8:00 to 10:00 pm shares. For $k \in \{pk, op\}$, if $q^{k,in}_{it}$ is observed we can calculate the share of category k calling opportunities that are in-network as

$$r^{k}_{it,s} = \frac{q^{k,in}_{it} / \hat{q}(v^{k,in}_{it,s})}{q^{k,in}_{it} / \hat{q}(v^{k,in}_{it,s}) + q^{k,out}_{it} / \hat{q}(v^{k,out}_{it,s})}.$$

Because $r^{k}_{it,s}$ follows a censored-normal distribution, where the underlying normal distribution is defined by $e_{it}^{r,k} + \alpha_{i,s}^{k}$, we can write the likelihood of $r_{it,s}^{k}$ as:

$$f^{r,k}(r_{it,s}^{k}|\alpha_{i,s}^{k},\Theta) = \begin{cases} \Phi(-\alpha_{i,s}^{k}/\sigma_e^k) & \text{if } r_{it,s}^{k} = 0 \\ \phi((r_{it,s}^{k} - \alpha_{i,s}^{k})/\sigma_e^k)/\sigma_e^k & \text{if } r_{it,s}^{k} \in (0,1) \\ 1 - \Phi((1-\alpha_{i,s}^{k})/\sigma_e^k) & \text{if } r_{it,s}^{k} = 1 \end{cases} .$$

When we only observe bounds on $q_{it}^{k,in}$ we can only calculate bounds for $r_{it,s}^{k}$: $\underline{r}_{it,s}^{k} = \underline{q}_{it}^{k,in}/(\theta_{it,s}^{k}\hat{q}(v_{it,s}^{*,k,in}))$ and $\bar{r}_{it,s}^{k} = \bar{q}_{it}^{k,in}/(\theta_{it,s}^{k}\hat{q}(v_{it,s}^{*,k,in}))$. Denoting $I_{it}^{b,k} = 1$ in situations where bounds can only be derived in category k, part II of the usage likelihood can be written as

$$l^{r,k}(r_{it,s}^{k}, I_{it}^{b,k}, \underline{r}_{it,s}^{k}, \bar{r}_{it,s}^{k}|\alpha_{i,s}^{k},\Theta) = f^{r,k}(r_{it,s}^{k})^{(1-I_{it}^{b,k})}\left(\Phi\left(\frac{\bar{r}_{it,s}^{k} - \alpha_{i,s}^{k}}{\sigma_e^k}\right) - \Phi\left(\frac{\underline{r}_{it,s}^{k} - \alpha_{i,s}^{k}}{\sigma_e^k}\right)\right)^{I_{it}^{b,k}}, \tag{24}$$

when $\underline{r}_{it,s}^{k} > 0$ and $\bar{r}_{it,s}^{k} < 1$. If $\underline{r}_{it,s}^{k} = 0$ and $\bar{r}_{it,s}^{k} < 1$ this share becomes

$$l^{r,k}(r_{it,s}^{k}, I_{it}^{b,k}, \underline{r}_{it,s}^{k}, \bar{r}_{it,s}^{k}|\alpha_{i,s}^{k},\Theta) = f^{r,k}(r_{it,s}^{k})^{(1-I_{it}^{b,k})}\left(\Phi\left(\frac{\bar{r}_{it,s}^{k} - \alpha_{i,s}^{k}}{\sigma_e^k}\right)\right)^{I_{it}^{b,k}}. \tag{25}$$

The likelihood when $\underline{r}_{it,s}^{k} \geq 0$ and $\bar{r}_{it,s}^{k} = 1$ is more complicated, and we defer discussion of that case below.

Calculation of the likelihood for 8:00 pm to 10:00 pm shares is similar to that of in-network and out-of-network shares. Denoting q_{it}^{9k} as observed 8:00 - 10:00 pm usage for $k \in \{pk, op\}$, the taste for 8:00 to 10:00 usage is simply $r_{it}^{9k} = q_{it}^{9k}/q_{it}^{k,out}$. Note that if $q_{it}^{k,out}$ is zero, we have no information about r_{it}^{9k}, so r_{it}^{9k} can only be computed when $q_{it}^{k,out} > 0$. If $q_{it}^{k,out}$ is observed exactly (rather than bounded), then the likelihood of r_{it}^{9k} is censored normal:

$$f^{9k}(r_{it}^{9k}|\alpha_{i,s}^{9k},\Theta) = \begin{cases} \Phi(-\alpha_{i,s}^{9k}/\sigma_e^{9k}) & \text{if } r_{it}^{k,9} = 0 \\ \phi((r_{it}^{9k} - \alpha_{i,s}^{k,9})/\sigma_e^{9k})/\sigma_e^{9k} & \text{if } r_{it}^{9k} \in (0,1) \\ 1 - \Phi((1-\alpha_{i,s}^{9k})/\sigma_e^{9k}) & \text{if } r_{it}^{9k} = 1 \end{cases} .$$

If we only have bounds on $q_{it}^{k,out}$, then we can also only put bounds on r_{it}^{9k},

$$r_{it}^{9k} \in \left[\frac{q_{it}^{9k}}{\bar{q}_{it}^{k,out}}, \frac{q_{it}^{9k}}{\underline{q}_{it}^{k,out}}\right] = \left[\underline{r}_{it}^{9k}, \bar{r}_{it}^{9k}\right],$$

and compute the probability of r_{it}^{9k} being in these bounds using the censored normal distribution.[54] We denote the probability of being in these bounds as $p^{9k}(\underline{r}_{it}^{9k}, \bar{r}_{it}^{9k}|\alpha_{i,s}^{9k}, \Theta)$.

If $\bar{r}_{it,s}^{k} = 1$ for $k \in \{pk, op\}$ then the likelihood is slightly more complicated. This stems from the way that bounds on in-network and out-of-network calls are constructed. The lower bound on out-of-network calls is the total number of outgoing calls to out-of-network numbers, plus the total number of non-free minutes used after an overage occurs. If this lower bound is zero, then total outgoing-calls to landlines were also zero. Total landline calls could be zero for two reasons: $r_{it,s}^{k} = 1$, or $r_{it}^{9k} = 0$. If the upper bound on $r_{it,s}^{k}$ binds, then r_{it}^{9k} could take any value. However, if the upper bound on $r_{it,s}^{k}$ does not bind, then r_{it}^{9k} must be zero. Following this logic, the joint likelihood of $\underline{r}_{it}^{k} \leq r_{it}^{k} \leq 1$ and $r_{it}^{9k} \in [0,1]$ is

$$p^{9k,0}(\underline{r}_{it,s}^{k}|\alpha_{i,s}^{k}, \alpha_{i,s}^{9k}, \Theta) = 1-\Phi((1-\alpha_{i,s}^{k})/\sigma_e^{k})+\Phi(-\alpha_{i,s}^{9k}/\sigma_e^{9k})\left[\Phi((1-\alpha_{i,s}^{k})/\sigma_e^{k}) - \Phi((\underline{r}_{it}^{k} - \alpha_{i,s}^{k})/\sigma_e^{k})\right].$$

Note that if $\underline{r}_{it,s}^{k} = 0$ the term $\Phi((\bar{r}_{it,s}^{k} - \alpha_{i,s}^{k})/\sigma_e^{k})$ is replaced with 0 in the likelihood. Denoting $\bar{I}_{it}^{b,k}$ as an indicator for $\bar{r}_{it,s}^{k} = 1$ the overall likelihood of 8:00 to 10:00 pm usage is

$$\begin{aligned} l^{9k}(r_{it}^{9k}, \underline{r}_{it}^{9k}, \bar{r}_{it}^{9k}, \underline{r}_{it,s}^{k}, I_{it}^{b,k}, \bar{I}_{it}^{b,k}|\alpha_{i,s}^{k}, \alpha_{i,s}^{k,9}, \Theta) \quad = \quad & (f^{9k}(r_{it}^{9k}|\alpha_{i,s}^{k}, \Theta))^{(1-I_{it}^{b,k})(1-\bar{I}_{it}^{b,k})} \\ & (p^{9k}(\underline{r}_{it}^{9k}, \bar{r}_{it}^{9k}|\alpha_{i,s}^{9k}, \Theta))^{(I_{it}^{b,k})(1-\bar{I}_{it}^{b,k})} \\ & (p^{9k,0}(\underline{r}_{it,s}^{k}|\alpha_{i,s}^{k}, \alpha_{i,s}^{9k}, \Theta))^{(I_{it}^{b,k})(\bar{I}_{it}^{b,k})}. \end{aligned} \qquad (26)$$

The joint likelihood of $\tilde{\boldsymbol{\theta}}_{it}$ and r_{it}^{k} will be the product of the l^{θ}, $l^{r,k}$ and $l^{9,k}$ for $k \in \{pk, op\}$. This likelihood is not equal to $f_{\boldsymbol{q}}$, however, because $\boldsymbol{q}_{it}$ is a function of $\tilde{\boldsymbol{\theta}}_{it}$ and r_{it}^{k} and we need to make the change of variables between them. Summarizing what we've outlined above, the transformation from the data to θ^{k},r^{k},$r^{9,k}$ is

$$\begin{aligned} \theta^{k} &= \frac{q^{out,k}}{\hat{q}(v^{*,out,k})} + \frac{q^{in,k}}{\hat{q}(v^{*,in,k})} \\ r^{k} &= \frac{q^{out,k}}{q^{out,k} + q^{in,k}\frac{\hat{q}(v^{*,out,k})}{\hat{q}(v^{*,in,k})}} \\ r^{9,k} &= \frac{q^{9,k}}{q^{out,k}}. \end{aligned}$$

We need to take the Jacobian determinant of this transformation and multiply it by each likelihood observation. We note that the Jacobian we take depends on what we observe. Below,

---

[54] Note that $q_{it}^{9k} < \underline{q}_{it}^{k,out}$, so $q_{it}^{9k} > 0$ implies the bounds on r_{it}^{9k} are within $[0,1]$.

we describe the case where all the six different q's are observed. The Jacobian is simpler if less data is observed. For example, if only θ^k were observed, and we could not compute the r's, we would only take the Jacobian of the transformation between θ^k_{it} and q^k_{it} for $k \in \{pk, op\}$. Defining $\boldsymbol{y} = (q^{pk,out}, q^{pk,in}, q^{9,pk}, q^{op,out}, q^{op,in}, q^{9,op})$ and $\boldsymbol{x} = (\theta^{pk}, r^{pk}, r^{9,pk}, \theta^{op}, r^{op}, r^{9,op})$,

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) \left| \det \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} \right) \right|.$$

For $k \in \{pk, op\}$, the derivatives we need are: $\partial\theta^k/\partial q^{k,out} = 1/\hat{q}(v^{k,out})$, $\partial\theta^k/\partial q^{k,in} = 1/\hat{q}(v^{k,in})$, $\partial r^{9,k}/\partial q^{k,out} = -q^{9,k}(q^{k,out})^{-2}$, $\partial r^{9,k}/\partial q^{9,k} = 1/q^{k,out}$,

$$\frac{\partial \theta^k}{\partial q^{9,k}} = \frac{\partial r^k}{\partial q^{9,k}} = \frac{\partial r^{9,k}}{\partial q^{k,in}} = 0,$$

and

$$\begin{aligned}
\frac{\partial r^k}{\partial q^{k,out}} &= q^{k,in}\frac{\hat{q}(v^{k,out})}{\hat{q}(v^{k,in})}\left(q^{k,out} + q^{k,in}\frac{\hat{q}(v^{k,out})}{\hat{q}(v^{k,in})}\right)^{-2} \\
\frac{\partial r^k}{\partial q^{k,in}} &= -q^{k,out}\frac{\hat{q}(v^{k,out})}{\hat{q}(v^{k,in})}\left(q^{k,out} + q^{k,in}\frac{\hat{q}(v^{k,out})}{\hat{q}(v^{k,in})}\right)^{-2}.
\end{aligned}$$

Then the Jacobian of the transformation that maps $\boldsymbol{y}$ into $\boldsymbol{x}$ is:

$$\begin{bmatrix}
\frac{\partial\theta^{pk}}{\partial q^{pk,out}} & \frac{\partial\theta^{pk}}{\partial q^{pk,in}} & 0 & 0 & 0 & 0 \\
\frac{\partial r^{pk}}{\partial q^{pk,out}} & \frac{\partial r^{pk}}{\partial q^{pk,in}} & 0 & 0 & 0 & 0 \\
\frac{\partial r^{9pk}}{\partial q^{pk,out}} & 0 & \frac{\partial r^{9pk}}{\partial q^{9pk}} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{\partial\theta^{op}}{\partial q^{op,out}} & \frac{\partial\theta^{op}}{\partial q^{op,in}} & 0 \\
0 & 0 & 0 & \frac{\partial r^{op}}{\partial q^{op,out}} & \frac{\partial r^{op}}{\partial q^{op,in}} & 0 \\
0 & 0 & 0 & \frac{\partial r^{9op}}{\partial q^{op,out}} & 0 & \frac{\partial r^{9op}}{\partial q^{9op}}
\end{bmatrix}.$$

The determinant of this will be

$$\det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}}\right) = \prod_{k \in pk,op}\left(\frac{\partial\theta^k}{\partial q^{k,out}}\frac{\partial r^k}{\partial q^{k,in}}\frac{\partial r^{9,k}}{\partial q^{9,k}} - \frac{\partial\theta^k}{\partial q^{k,in}}\frac{\partial r^k}{\partial q^{k,out}}\frac{\partial r^{9,k}}{\partial q^{9,k}}\right).$$

We conclude this section with some notes on computing the consumer expected utility, which we need to compute the choice probabilities. The utilities $U_{itj,s}$ are defined in equation (17) for all plans in consumers' choice sets. These depend on plan-specific calling threshold vectors $\mathbf{v}^*_{itj,s}$. Those for the chosen plan have already been computed, and those for non-chosen plans can be computed given the calculated beliefs. Next, notice that

$$V\left(q(v_{itj}^k, x_{it}^k), x_{it}^k\right) = x_{it}^k \frac{1}{\beta}\left(\ln\left(\hat{q}(v_{itj}^k)\right) - \hat{q}(v_{itj}^k)\right) \tag{27}$$

is linear in x_{it}^k for $k \in$ {pk-out, pk-in, op-out, op-in} and hence $\sum_k V(q_{it}^k, x_{it}^k)$ is linear in θ_{it}^{pk} and θ_{it}^{op}. Thus $E\left[\sum_k V(q_{it}^k, x_{it}^k)\right]$ can be computed analytically (up to evaluation of the standard normal cumulative distribution). Moreover, the expected price $E[P(q)]$ is a function of the expected amount θ_{it}^{pk} exceeds $Q_{ijt}/\hat{q}(v_{itj,s}^{pk})$ (or $x_{it}^{pk,out}$ exceeds $Q_{ijt}/\hat{q}(v_{itj,s}^{pk,out})$ for free-in-network), which we can also evaluate analytically in all cases except for when plan 2 offers free in-network minutes.[55]

E Estimation Details

We wrote the program to evaluate the likelihood in R and Fortran. The evaluation of this likelihood is computationally intensive for two reasons: first, it must be evaluated at many simulation draws; second, for each choice a consumer could make, at each time period and each draw, we often must solve for $\mathbf{v}_{it}^*$ and $\alpha_i^{9,op}$ using a nonlinear equation solver. Our estimation method therefore falls into an inner-loop outer-loop framework, where the inner loop is the solution of the $\mathbf{v}_{it}^*$'s and $\alpha_i^{9,op}$'s, and the outer loop maximizes the likelihood.

We summarize the algorithm for computing these variables in four steps. Step 1 is to compute $\alpha_{i,s}^{9op}$ conditional on the simulated draws and the other model parameters. Recall that we assume that a consumer's average taste for weekday-evening landline-usage is the same thirty minutes before and after 9pm. For each consumer i and each simulation draw s, we compute $\alpha_{i,s}^{9op}$ as the solution to equation (20) in Appendix C, which extends equation (7) to account for in-network calling. As this equation does not have an analytic solution, we compute $\alpha_{i,s}^{9op}$ with a nonlinear equation solver. The result of this step is used to compute the structural error for r_{it}^{9op}.

The next three steps compute the calling threshold vector $\mathbf{v}_{it,s}^*$ and $\tilde{\boldsymbol{\theta}}_{it,s}$ period-by-period. Because the $\mathbf{v}_{it,s}^*$ is a function of past values of $\tilde{\boldsymbol{\theta}}_{it,s}$ through the Bayesian learning and the AR1 process, these three steps are iterated across both individuals i, and time periods t. Step 2 calculates consumer beliefs about $\tilde{\boldsymbol{\theta}}_{it,s}$ in two parts following Section 4.4. First, consumer beliefs about μ_i^{pk}, $(\tilde{\mu}_{it,s}^{pk}, \tilde{\boldsymbol{\sigma}}_{it}^{-2})$ are updated via Bayes rule by conditioning on the lagged value $\tilde{\theta}_{i,t-1,s}^{pk}$. Second, beliefs about $\tilde{\boldsymbol{\theta}}_{it,s}$ are computed from $(\tilde{\mu}_{it,s}^{pk}, \tilde{\boldsymbol{\sigma}}_{it}^{-2})$, $\mu_{i,s}^{op}$ and the the lagged value $\tilde{\boldsymbol{\theta}}_{i,t-1,s}$ which enters through the AR1 process. (No updating is required for $t = 1$.) In step 3 we calculate $\mathbf{v}_{it,s}^*$ following it's characterization in Appendix C, which depends on the beliefs calculated in step 2. Recall that

[55] In this case we again approximate the probability with Gaussian quadrature.

components of $\mathbf{v}^*_{it}$ are either known to be 0 cents or 11 cents or must be calculated by numerically solving a first-order condition (either equation (18) or (19) which are the extensions to equation (2) that account for in-network calling given in Appendix C). In step 4, we calculate $\tilde{\boldsymbol{\theta}}_{it,s}$. When $\boldsymbol{\theta}_{it,s}$ is not censored, we can compute $\tilde{\boldsymbol{\theta}}_{it,s}$ from observed usage conditional on β and $\mathbf{v}^*_{it,s}$ using equations (21)-(22) in Appendix C. When censoring occurs, we use the simulated value for $\tilde{\boldsymbol{\theta}}_{it,s}$.

With $\alpha^{9op}_{i,s}$, $\tilde{\boldsymbol{\theta}}_{it,s}$ and $\mathbf{v}^*_{it,s}$ in hand we can compute the choice probabilities and the density of observed usage in equation (10). We optimize our likelihood in two steps. The first step uses a Nelder-Mead optimizer to get close to the optimum. From there we use a Newton-Raphson optimizer to reach the optimum within a tighter tolerance. Because the optimization algorithms will stop at local optima, it is important to have good starting points. To arrive at starting points for the model, we choose the usage parameters (the means and variances of the μ's, α's, and ε's) and the β to match observed usage.[56] Conditional on these choices of usage parameters, we choose initial belief parameters to match the observed plan shares. To do this, we use our model to simulate plan shares for the 2002 to 2003 school year and the 2003 to 2004 school year, and match those simulated shares to the observed shares during these two years. We chose to split the data in that way to exploit the fact that plan 0 stopped offering free off-peak minutes at the beginning of the 2003 to 2004 school year.

F Nested-Logit Calibration Exercise

As discussed in Section 8, to run our endogenous-price counter-factual simulations we first calibrate the inclusive-value parameter λ of a nested-logit specification. Our algorithm for doing this is as follows: First, we limit the analysis to publicly available prices, which means that we focus on the price offerings of the three firms AT&T, Cingular, and Verizon.[57] Second, we draw 1000 simulated consumers and simulate their usage and plan choices for one year. We assume that the three firms each offer menus of four three-part tariffs and play a symmetric Nash equilibrium. Finally, assuming that marginal costs are zero, we solve for equilibrium prices and compare them to prices observed over the two year period.[58] Formally, firm j offers $k = 1, .., 4$ plans, where the

[56] We assume that v^* is equal to 3 cents for plan 3, 5 cents for plan 2, and 8 cents for plan 1, and maximize the likelihood of usage conditional on those guesses at v^*. We chose those values of v^* because they matched the average values of v^* that were produced by simulating the model at parameters which were in the neighborhood of the estimates. We stress that we only use the guesses at v^* to arrive at starting points; we solve for the endogenous v^*'s when running the full simulated maximum likelihood.

[57] Sprint did not offer any local plans, so we did not include them.

[58] We restrict included minute allowances, Q_j, to weakly increase with fixed fees, M_j, and bound overage rates $p_j \leq 50$¢ (see footnote 46).

plans are ordered according to their fixed fees, and we denote the predicted price schedule as a function of λ as $(\hat{M}_{jk}(\lambda), \hat{p}_{jk}(\lambda), \hat{Q}_{jk}(\lambda))$ and the observed price schedule of a firm during month t as $(M_{jkt}, p_{jkt}, Q_{jkt})$. We choose λ to minimize the squared difference between observed and predicted prices:

$$\hat{\lambda} = \arg\min_{\lambda} \sum_{j,k,t} \left((\hat{M}_{jk}(\lambda) - M_{jkt})^2 + (\hat{p}_{jk}(\lambda) - p_{jkt})^2 + (\hat{Q}_{jk}(\lambda) - Q_{jkt})^2 \right).$$

This calibration exercise yields $\hat{\lambda} = 0.20$. We used a grid search to find the optimal $\hat{\lambda}$.

Table 8: Parameter Robustness to λ in Nested Logit Specification

Coefficient	$\lambda = 1$	$\lambda = 0.2$	Coefficient	$\lambda = 1$	$\lambda = 0.2$
β	4.024	4.333	μ_α^{9pk}	-0.004	-0.006
$\tilde{\mu}_0^{pk}$	-19.275	-20.302	$(\sigma_\alpha^{9pk})^2$	0.06	0.06
μ_0^{pk}	103.327	104.975	$(\sigma_e^{9pk})^2$	0.104	0.104
μ_0^{op}	97.512	96.272	$(\sigma_e^{9op})^2$	0.116	0.116
$\tilde{\sigma}_{\mu^{pk}}$	143.196	143.434	φ	0.579	0.58
$\tilde{\sigma}_1$	12.876	12.756	α	0.096	0.097
$\sigma_{\mu^{pk}}$	78.584	78.599	Price Consideration	0.063	0.063
$\sigma_{\mu^{op}}$	161.784	161.053	Outside Good Utility	-71.191	-71.926
ψ^{pk}	-0.043	-0.047	δ_r	0.003	0.003
ψ^{op}	0.228	0.223	μ_α^{pk}	0.35	0.347
ρ_μ	0.981	0.981	μ_α^{op}	0.401	0.397
$\tilde{\sigma}_\varepsilon^{pk}$	163.574	177.984	$(\sigma_\alpha^{pk})^2$	0.035	0.035
σ_ε^{pk}	182.521	183.184	$(\sigma_\alpha^{op})^2$	0.039	0.039
ρ_ε	0.407	0.41	$(\sigma_e^{pk})^2$	0.03	0.03
σ_ε^{op}	306.574	305.423	$(\sigma_e^{op})^2$	0.025	0.025

In Table 8, we show how our parameter estimates from our primary specification ($\lambda = 1$) differ when we re-estimate the model for λ equal to 0.2. The table shows that parameter estimates are relatively insensitive to the chosen value of λ. The largest changes are in the estimates of β and $\tilde{\sigma}_\varepsilon^{pk}$. When λ moves from 1 to 0.2, β rises from 4 to 4.3. However, this corresponds to only a small change in price sensitivity: A price increase from 0 to 11 cents decreases usage by 31% at our estimates but by 32% with $\beta = 4.33$. The $\tilde{\sigma}_\varepsilon^{pk}$ rises from 164 to 178 minutes, which reflects a modest decrease in the amount of volatility bias (δ_ε rises from 0.89 to 0.97). Again, this change will not have a big impact on any inferences we make with our model; most of the impact of biases happens through overconfidence and mean biases. The other parameter estimates show even less sensitivity to λ.