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

Michael D. Grubb and Matthew Osborne

Online Appendix

May 2014

A Data Details

A.1 University Price Data

Table 6: 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 6. This price series was inferred from billing data rather than directly observed. 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 able to reliably infer this pricing information for popular plans from August 2002 to July 2004. We exclude other dates and the 11 percent of bills from unpopular plans (national plans, free-long-distance plans, and expensive local plans), grouping them with the outside option in our structural model. In fact, we treat switching to an unpopular plan the same as quitting service, hence we also drop all remaining bills once a customer switches to an unpopular plan, even if they eventually switch back to a popular plan.

As stated in Section II.A, an additional 7 percent of individuals are excluded from our structural estimation due to data problems. In particular, we exclude individuals with substantially negative bills, indicating either billing errors or ex post renegotiated refunds that are outside our model. Also excluded are individuals who have infeasible choices recorded (plans outside the choice set or negative in-network calling) and 8 individuals for whom we could not find starting points (initial parameter values from which to begin maximizing the likelihood) with positive likelihood.

A.2 Public Price Data

Table 7 shows a subset of data obtained from EconOne: publicly available local calling plans for October 2003 in the same geographic market as the university. (Sprint did not offer local plans.)

Table 7: Publicly Available Local Calling Plans - October 2003

	AT&T			Cingular			Verizon		
Plan	M	Q	p	M	Q	p	M	Q	p
1	29.99	350	0.45	29.99	300	0.49	29.99	300	0.45
2	39.99	600	0.40	39.99	600	0.49	39.99	500	0.45
3	49.99	800	0.40	49.99	1000	0.45	49.99	700	0.40
4	59.99	1050	0.35	69.99	1200	0.45	59.99	1000	0.40

A.3 Additional Evidence of Inattention

If consumers are attentive to the remaining balance of included minutes during the billing cycle they should use this information to continually update their beliefs about the likelihood of an overage and a high marginal price ex post. Following an optimal dynamic program, an attentive consumer should (all else equal) reduce her usage later in the month following unexpectedly high usage earlier in the month. This prediction should be true for any consumers who are initially uncertain whether they will have an overage in the current month. For these consumers, the high usage shock early in the month increases the likelihood of an overage, thereby increasing their expected ex post marginal price, and causing them to be more selective about calls. If calling opportunities arrived independently throughout the month, this strategic behavior by the consumer would lead to negative correlation between early and late usage within a billing period. However, looking for negative correlation in usage within the billing period is a poor test for this dynamic behavior because it is likely to be overwhelmed by positive serial correlation in taste shocks.

To test for dynamic behavior by consumers within the billing period, we use our data set of individual calls to construct both fortnightly and weekly measures of peak usage. A simple regression of usage on individual fixed effects and lagged usage shows strong positive serial correlation. However, we take advantage of the following difference: Positive serial correlation between taste shocks in periods t and $(t-1)$ should be independent of whether periods t and $(t-1)$ are in the same or adjacent billing cycles. However, following unexpectedly high usage in period $(t-1)$, consumers should cut back usage more in period t if the two periods are in the same billing cycle. Thus by including an interaction effect between lagged usage and an indicator for the lag being in the same billing cycle as the current period, we can separate strategic behavior within the month from serial correlation in taste shocks.

Table 8 shows a regression of log usage on lagged usage and the interaction between lagged usage and an indicator equal to 1 if period $(t-1)$ is in the same billing cycle as period t. We also include time and individual fixed effects and correct for bias induced by including both individual fixed effects and lags of the dependent variable in a wide but short panel (Roodman, 2009). Reported analysis is for plan 1, the most popular three-part tariff. As expected, positive serial correlation in demand shocks leads to a positive and significant coefficient on lagged usage in the full sample (column 1) and most subsamples (columns 2-6). If consumers adjust their behavior dynamically within the billing cycle in response to usage shocks, then we expect the interaction effect to be negative. In the full sample (column 1) the interaction effect has a positive point estimate, but is not significantly different from zero. This result suggests that consumers are not attentive to past usage during the course of the month.

Table 8: Dynamic Usage Pattern at Fortnightly Level.

	(1)	(2)	(3)	(4)	(5)	(6)
Overage Percentage	0-100%	0	1-29%	30-70%	71-99%	100%
$\ln(q_{t-1})$	0.649***	0.609***	0.522***	0.514***	-1.046	0.958***
	(0.0270)	(0.0524)	(0.0437)	(0.0695)	(1.060)	(0.0441)
SameBill*$\ln(q_{t-1})$	0.00692	0.0150	0.0256	-0.0222	-0.0837	3.685
	(0.0108)	(0.0197)	(0.0182)	(0.0229)	(1.174)	(4.745)
Observations	9062	3717	3222	1830	217	76
Number of individuals	385	166	130	87	11	6

Dependent variable $\ln(q_t)$. Standard errors in parentheses. Time and individual fixed effects.
Key: *** p<0.01, ** p<0.05, * p<0.1

Consumers who either never have an overage (43 percent of plan 1 subscribers) or always have an overage (3 percent of plan 1 subscribers) should be relatively certain what their ex post marginal price will be, and need not adjust calling behavior during the month. For instance, consumers who always make overages may only make calls worth more than the overage rate throughout the month. For such consumers we would expect to find no interaction effect, and this may drive the result when all consumers are pooled together as in our first specification. As a result, we divide consumers into groups by the fraction of times within their tenure that they have overages. We repeat our first specification for different *overage-risk groups* in Columns 2-6 of Table 8. The interaction effect is indistinguishable from zero in all overage risk groups. Moreover, in unreported analysis, more flexible specifications that include nonlinear terms[1] and a similar analysis at the weekly rather than fortnightly level all estimate an interaction effect indistinguishable from zero. There is simply no evidence that we can find that consumers strategically cut back usage at the end of the month following unexpectedly high initial usage. We conclude that consumers are inattentive to their remaining balance of included minutes during the billing cycle.

A.4 Plan Switching Calculations

We make two calculations for each switch from an existing plan j to an alternate plan j' that cannot be explained by a price cut for plan j'. First, we calculate how much the customer would have saved had they signed up for the new plan j' initially, holding their usage from the original plan j fixed. By this calculation, average savings are \$10.87 to \$15.24 per month and 60 to 61 percent of switches save consumers money. We calculate bounds because we cannot always distinguish in-network and out-of-network calls. Average savings are statistically greater than zero at the 99 percent level. The 60-61 percent rates of switching in the "right" direction are statistically greater than 50 percent at the 95 percent level. This calculation is based on 99 of the 136 switches which cannot be explained by price decreases. The remaining 37 switches occur so soon after the customer joins that there is no usage data prior to the switch that is not from a pro-rated bill.

Second, we calculate how much money the customer would have lost had they remained on existing plan j rather than switching to the new plan j', now holding usage from plan j' fixed. By this calculation, average savings are \$13.80 to \$24.56 per month, and 61 to 69 percent of switches save money. This calculation is based on 132 of the 137 switches which can not be explained by

[1] Average q_t will vary with expected marginal price, which is proportional to the probability of an overage. The probability of an overage in a billing period which includes periods t and $(t-1)$ increases nonlinearly in q_{t-1}. In one specification, we first fit a probit on the likelihood of an overage as a function of the first fortnights usage, and then used the estimated coefficients to generate overage probability estimates for all fortnights. We then included these (lagged) values as explanatory variables. In an alternative specification we added polynomial terms of lagged q_{t-1}.

price decreases. The calculation cannot be made for the remaining 5 switches since there is no usage data following the switch that is not from a pro-rated bill. Figures are significant at the 95 percent or 99 percent confidence level.

As explained in Section II.B, the first and second calculations described above provide lower and upper bounds, respectively, on the benefits of switching plans. We therefore conclude that consumers' expected benefit from switching is between \$10.87 and \$24.56 per month and 60 to 69 percent of switches save money.

B Model Details

A Model Guide: Tables 9 - 12 provide a guide to model parameters.

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 $V_q(q_{it}^k,\theta_{it}^k) = \left(1 - q_{it}^k\theta_{it}^k\right)/\beta$ 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_{itj}^{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. (Notice that substituting $q\left(v^*,\theta_{it}^k\right) = \theta_{it}^k\left(1-\beta v^*\right)$

Table 9: Model Guide: Payments, Preferences, and Usage

Parameter	Description	Distribution or equation
$q_{itj}^{billable}$	billable minutes	$q_{itj}^{billable} = q_{it}^{pk} + OP_j q_{it}^{op}$
OP_j	indicator for costly off-peak plan j	constant
P_j	price plan j	$P_j(\mathbf{q}_{it}) = M_j + p_j \max\{0, q_{itj}^{billable} - Q_j\}$
M_j	monthly fee plan j	constant
Q_j	allowance plan j	constant
p_j	overage rate plan j	constant
V	value function	$V\left(q_{it}^k, \theta_{it}^k\right) = \frac{1}{\beta} q_{it}^k \left(1 - \frac{1}{2}\left(q_{it}^k / \theta_{it}^k\right)\right)$
u_{itj}	utility function	$u_{itj} = \sum_{k\in\{pk,op\}} V\left(q_{it}^k, \theta_{it}^k\right) - P_j(\mathbf{q}_{it}) + \eta_{itf}$
U_{itj}	expected utility	equation (5)
β	price sensitivity	constant
O	outside good utility	constant
P_C	plan consideration Pr.	constant
η_{itf}	logit error	iid logit
θ_{it}^k	# calling opportunities	$\theta_{it}^k = \max\{0, \tilde{\theta}_{it}^k\}$
$\tilde{\theta}_{it}^k$	latent taste shock	$\tilde{\boldsymbol{\theta}}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{it}$
v_{itj}^k	calling threshold	equation (4)
$\mathbf{v}_{itj}^*$	calling threshold vector	$\mathbf{v}_{itj}^* = (v_{itj}^{pk}, v_{itj}^{op})$
$\hat{q}(v_{itj}^k)$	Frac. calling opp. value $> v_{itj}^k$	$\hat{q}(v_{itj}^k) = (1 - \beta v_{itj}^k)$
q_{it}^k	minutes of calling (usage)	$q_{it}^k = \theta_{it}^k \hat{q}(v_{itj}^k)$

Subscripts and superscripts denote customer i, time t, plan j, firm f, and calling category $k \in \{\text{pk, op}\}$.

Table 10: Model Guide: Tastes & Signals

Parameter	Description	Distribution
$\begin{bmatrix} \tilde{\mu}_{i1}^{pk} \\ \mu_i^{pk} \\ \mu_i^{op} \end{bmatrix}$	$\begin{bmatrix} \text{Point Estimate } \mu_i^{pk} \\ \text{Peak Type} \\ \text{Off-Peak Type} \end{bmatrix}$	$N\left(\begin{bmatrix} \tilde{\mu}_0^{pk} \\ \mu_0^{pk} \\ \mu_0^{op} \end{bmatrix}, \begin{bmatrix} (\tilde{\sigma}_\mu^{pk})^2 & \rho_{\tilde{\mu},pk}\sigma_\mu^{pk}\tilde{\sigma}_\mu^{pk} & \rho_{\tilde{\mu},op}\sigma_\mu^{op}\tilde{\sigma}_\mu^{pk} \\ \rho_{\tilde{\mu},pk}\sigma_\mu^{pk}\tilde{\sigma}_\mu^{pk} & (\sigma_\mu^{pk})^2 & \rho_\mu\sigma_\mu^{pk}\sigma_\mu^{op} \\ \rho_{\tilde{\mu},op}\sigma_\mu^{op}\tilde{\sigma}_\mu^{pk} & \rho_\mu\sigma_\mu^{pk}\sigma_\mu^{op} & (\sigma_\mu^{op})^2 \end{bmatrix}\right)$
$\begin{bmatrix} s_{it} \\ \varepsilon_{it}^{pk} \\ \varepsilon_{it}^{op} \end{bmatrix}$	$\begin{bmatrix} \text{Signal} \\ \text{Peak Shock} \\ \text{Off-Peak Shock} \end{bmatrix}$	$N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{s,pk}\sigma_\varepsilon^{pk} & \rho_{s,op}\sigma_\varepsilon^{op} \\ \rho_{s,pk}\sigma_\varepsilon^{pk} & (\sigma_\varepsilon^{pk})^2 & \rho_\varepsilon\sigma_\varepsilon^{pk}\sigma_\varepsilon^{op} \\ \rho_{s,op}\sigma_\varepsilon^{op} & \rho_\varepsilon\sigma_\varepsilon^{pk}\sigma_\varepsilon^{op} & (\sigma_\varepsilon^{op})^2 \end{bmatrix}\right)$

Table 11: Model Guide: Beliefs and Biases

Parameter	Description	Equation
δ	overconfidence	
$\tilde{\mu}_{it}^{pk}$	updated point est.	Bayes rule, equation (17)
σ_t	$SD[\mu_i^{pk} \mid \tilde{\mu}_{it}^{pk}]$	Bayes rule ($t > 1$) equation (18)
$\tilde{\sigma}_t$	$\widetilde{SD}[\mu_i^{pk} \mid \tilde{\mu}_{it}^{pk}]$	$\tilde{\sigma}_t = \delta\sigma_t$
$\tilde{\mu}_{\theta it}^{pk}$	$\tilde{E}\left[\theta_{it}^{pk} \mid s_{it}\right]$	equation (6)
$\sigma_{\theta t}$	$SD\left[\theta_{it}^{pk} \mid s_{it}\right]$	equation (7)
$\tilde{\sigma}_{\theta t}$	$\widetilde{SD}\left[\theta_{it}^{pk} \mid s_{it}\right]$	$\tilde{\sigma}_{\theta t} = \delta\sigma_{\theta t}$
b_1	aggregate mean bias	$b_1 = \tilde{\mu}_0^{pk} - \mu_0^{pk}$
b_2	conditional mean bias	$b_2 = 1 - \frac{Cov(\mu_i^{pk}, \tilde{\mu}_{i1}^{pk})}{Var(\tilde{\mu}_{i1}^{pk})} = 1 - \rho_{\tilde{\mu},pk}\frac{\sigma_\mu^{pk}}{\tilde{\sigma}_\mu^{pk}}$. equation (8)

Operators E and SD denote population moments, while $\tilde{E}$ and $\widetilde{SD}$ are with respect to individual i's beliefs.

Table 12: Model Guide: Near-9pm Calling Parameters

Parameter	Description	Distribution or equation
$r_{it}^{k,9}$	share of k-calling near 9pm	$\tilde{r}_{it}^{k,9}$ censored on $[0,1]$
$\tilde{r}_{it}^{k,9}$	latent shock	$\tilde{r}_{it}^{k,9} = \alpha_i^{k,9} + e_{it}^{k,9}$
$\alpha_i^{pk,9}$	pk-9pm type	$\alpha_i^{pk,9} \sim N(\mu_\alpha^{pk,9}, (\sigma_\alpha^{pk,9})^2)$
$e_{it}^{k,9}$	pk-9pm shock	$e_{it}^{k,9} \sim N(0, (\sigma_e^{k,9})^2)$, iid across i, t, and k.
$\alpha_i^{op,9}$	op-9pm type	Defined implicitly by equation (9)

Subscripts and superscripts denote customer i, time t, and calling category $k \in \{\text{pk, op}\}$.

from equation (12) into $V_q(q_{it}^k, \theta_{it}^k)$ yields v^*). Thus equation (13) 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 simplification yields equation (4). It is apparent by inspection that equation (4) has a unique solution.

Equation (4) may seem counter-intuitive, because the optimal v_{itj}^{pk} is greater than the expected marginal price, $p_j \Pr(q(v_{itj}^{pk}, \theta_{it}^{pk}) > Q_j \mid \Im_{it})$. This is because the reduction in consumption from raising v_{itj}^{pk} is proportional to θ_{it}^{pk}. Raising v_{itj}^{pk} cuts back on calls valued at v_{itj}^{pk} more heavily in high demand states when they cost p_j and less heavily in low demand states when they cost 0.

Figure 9 plots the optimal calling threshold v_{itj}^{pk} as a function of consumer beliefs and the corresponding plan choice. The distinct regions visible on the contour plot correspond to the plan choice regions identified in Figure 7. Figure 9 shows that, on three-part tariffs, v_{itj}^{pk} increases with the perceived mean and variance of calling opportunities because both increase the upper tail of usage and hence the likelihood of paying overage fees. Figure 9 also shows that consumers who choose a three-part tariff use calling thresholds that are small relative to the overage rates of \$0.35 or \$0.45 per minute. This is because if the perceived likelihood of an overage were large the consumer would have chosen a larger plan.

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

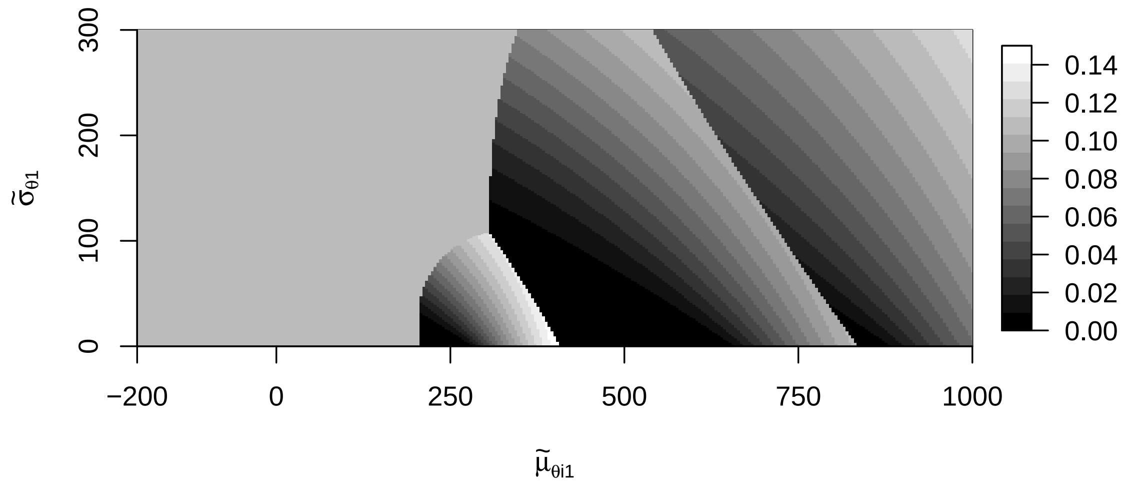

Figure 9: Calling threshold v_{itj}^{pk} as a function of initial beliefs $\{\tilde{\mu}_{\theta i1}, \tilde{\sigma}_{\theta 1}\}$ and the corresponding plan choice (shown in Figure 7) implied by the model evaluated at fall 2002 prices given $\beta = 2$.

calls above some threshold v^* but raise this calling threshold to the overage rate p upon receiving a bill-shock alert 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 alert. If $\theta > \theta^*$ then a bill-shock alert is received and Figure 10 illustrates demand curves and calling thresholds before and after the alert. Prior to receiving a bill-shock alert, the customer has θ^* calling opportunities. Of these, fraction $\hat{q}(v^*)$ are worth more than v^* so that prior to the alert the consumer makes $\theta^* \hat{q}(v^*) = Q$ calls with total value $V(Q, \theta^*)$ and declines $\theta^*(1 - \hat{q}(v^*))$ calling opportunities. After receiving a bill-shock warning, the customer has $\theta^a = \theta - \theta^*$ additional calling opportunities. Of these, fraction $\hat{q}(p)$ are worth more than the overage rate p so that after the alert the consumer makes an additional $\theta^a \hat{q}(p)$ calls with total value $V(q(p, \theta^a), \theta^a)$.

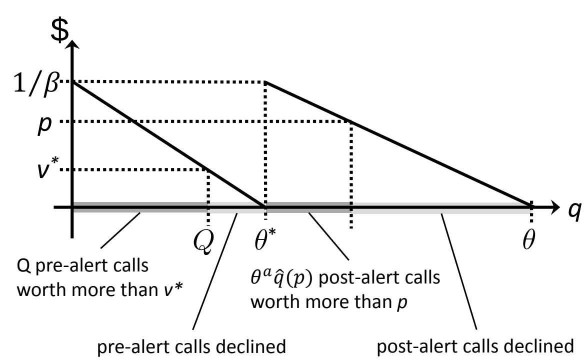

Figure 10: Inverse demand curves and calling thresholds before and after a bill-shock alert.

Notice that calls with the highest value $1/\beta$ are made both before and after the alert. This is because the demand curves' ordering of calls from high to low value is not chronological. Both low and high value calls arrive throughout the month.

We now characterize the optimal calling threshold v^* with bill-shock alerts. The consumer chooses v^* to maximize

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

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

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

The multiplicative demand assumption implies $\frac{d}{dv^*} q\left(v^*, \theta\right) = \theta \frac{d}{dv^*} \hat{q}\left(v^*\right)$ and $d\theta^*/dv^* = -\left(Q/\hat{q}^2\left(v^*\right)\right) \frac{d}{dv^*} \hat{q}\left(v^*\right)$. Therefore the $\frac{d}{dv^*} \hat{q}\left(v^*\right)$ terms cancel and the first-order condition is

$$v^* \int_0^{\theta^*} \theta f\left(\theta\right) d\theta = \frac{Q}{\hat{q}^2\left(v^*\right)} \int_{\theta^*}^{\infty} \left(V_\theta\left(Q, \theta^*\right) - V_\theta\left(q\left(p, \theta^a\right), \theta^a\right)\right) f\left(\theta\right) d\theta.$$

Notice that because $V_\theta\left(q, \theta\right) = q^2/\left(2\beta\theta^2\right)$ and $q\left(p, \theta\right) = \theta \hat{q}\left(p\right)$ it holds that $V_\theta\left(q\left(p, \theta\right), \theta\right) = \hat{q}^2\left(p\right)/2\beta$ is independent of θ. Substituting this expression in yields

$$v^* \int_0^{\theta^*} \theta f\left(\theta\right) d\theta = \frac{Q}{2\beta} \left(1 - \left(\frac{\hat{q}\left(p\right)}{\hat{q}\left(v^*\right)}\right)^2\right) \left(1 - F\left(\theta^*\right)\right).$$

Rearranging terms, this expression is equivalent to

$$v^* = \frac{Q}{2\beta} \left(1 - \left(\frac{\hat{q}\left(p\right)}{\hat{q}\left(v^*\right)}\right)^2\right) \frac{\left(1 - F\left(\theta^*\right)\right)}{\int_0^{\theta^*} \theta f\left(\theta\right) d\theta}, \tag{14}$$

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

B.3 Bayesian Updating

In this appendix we use operators E and V to denote mean and variance in the population or given objective probabilities. We use operators $\tilde{E}$ and $\tilde{V}$ to denote mean and variance given an individual consumers' subjective beliefs.

Updating from signal s_{it} At the beginning of each period, each consumer receives a signal s_{it} that is informative about the taste innovation $\boldsymbol{\varepsilon}_{it}$. Given the joint distribution of $\boldsymbol{\varepsilon}_{it}$ and s_{it} and the restriction $\rho_{s,op} = \rho_{s,pk}\rho_{\varepsilon}$, conditional on the signal s_{it}: $\boldsymbol{\varepsilon}_{it}$ is normally distributed with mean

$$E\left(\begin{bmatrix}\varepsilon_{it}^{pk} & \varepsilon_{it}^{op}\end{bmatrix} \mid s_{it}\right) = \begin{bmatrix}\rho_{s,pk}\sigma_{\varepsilon}^{pk}s_{it} & \rho_{s,pk}\rho_{\varepsilon}\sigma_{\varepsilon}^{op}s_{it}\end{bmatrix},$$

and variance

$$V\left(\begin{bmatrix}\varepsilon_{it}^{pk} & \varepsilon_{it}^{op}\end{bmatrix} \mid s_{it}\right) = \begin{bmatrix} (\sigma_{\varepsilon}^{pk})^2\left(1-\rho_{s,pk}^2\right) & \rho_{\varepsilon}\sigma_{\varepsilon}^{pk}\sigma_{\varepsilon}^{op}\left(1-\rho_{s,pk}^2\right) \\ \rho_{\varepsilon}\sigma_{\varepsilon}^{pk}\sigma_{\varepsilon}^{op}\left(1-\rho_{s,pk}^2\right) & (\sigma_{\varepsilon}^{op})^2\left(1-\rho_{s,pk}^2\rho_{\varepsilon}^2\right) \end{bmatrix}.$$

However, as consumers underestimate the unconditional standard deviations of s_{it} and ε_{it}^{pk} by a factor δ, consumers perceive the conditional distribution to be normal with mean

$$\tilde{E}\left(\begin{bmatrix}\varepsilon_{it}^{pk} & \varepsilon_{it}^{op}\end{bmatrix} \mid s_{it}\right) = \begin{bmatrix}\rho_{s,pk}\sigma_{\varepsilon}^{pk}s_{it} & \frac{1}{\delta}\rho_{s,pk}\rho_{\varepsilon}\sigma_{\varepsilon}^{op}s_{it}\end{bmatrix}, \tag{15}$$

and variance

$$\tilde{V}\left(\begin{bmatrix}\varepsilon_{it}^{pk} & \varepsilon_{it}^{op}\end{bmatrix} \mid s_{it}\right) = \begin{bmatrix} (\tilde{\sigma}_{\varepsilon}^{pk})^2\left(1-\rho_{s,pk}^2\right) & \rho_{\varepsilon}\tilde{\sigma}_{\varepsilon}^{pk}\sigma_{\varepsilon}^{op}\left(1-\rho_{s,pk}^2\right) \\ \rho_{\varepsilon}\tilde{\sigma}_{\varepsilon}^{pk}\sigma_{\varepsilon}^{op}\left(1-\rho_{s,pk}^2\right) & (\sigma_{\varepsilon}^{op})^2\left(1-\rho_{s,pk}^2\rho_{\varepsilon}^2\right) \end{bmatrix}, \tag{16}$$

where $\tilde{\sigma}_{\varepsilon}^{pk} = \delta\sigma_{\varepsilon}^{pk}$.[2]

The preceding beliefs about the taste innovation $\boldsymbol{\varepsilon}_{it}$ correspond to the belief that $\boldsymbol{\theta}_{it}$ is normally distributed with mean

$$\tilde{E}\left(\begin{bmatrix}\theta_{it}^{pk} \\ \theta_{it}^{op}\end{bmatrix} \mid s_{it}\right) = \begin{bmatrix}\tilde{\mu}_{it}^{pk} + \rho_{s,pk}\sigma_{\varepsilon}^{pk}s_{it} \\ \mu_{i}^{op} + \rho_{s,op}\sigma_{\varepsilon}^{op}s_{it}\end{bmatrix},$$

and variance

$$\tilde{V}\left(\begin{bmatrix}\theta_{it}^{pk} \\ \theta_{it}^{op}\end{bmatrix} \mid s_{it}\right) = \begin{bmatrix}\tilde{\sigma}_t^2 + (1-\rho_{s,pk}^2)(\tilde{\sigma}_{\varepsilon}^{pk})^2 & (1-\rho_{s,pk}^2)\rho_{\varepsilon}\tilde{\sigma}_{\varepsilon}^{pk}\sigma_{\varepsilon}^{op} \\ (1-\rho_{s,pk}^2)\rho_{\varepsilon}\tilde{\sigma}_{\varepsilon}^{pk}\sigma_{\varepsilon}^{op} & (1-\rho_{s,pk}^2\rho_{\varepsilon}^2)(\sigma_{\varepsilon}^{op})^2\end{bmatrix}.$$

[2] As beliefs about off-peak tastes are not identified by our data, we assume that consumers understand the unconditional variance of off-peak tastes $(\sigma_{\varepsilon}^{op})^2$. This assumption implies that consumers also correctly understand the conditional variance $(\sigma_{\varepsilon|s}^{op})^2 = (\sigma_{\varepsilon}^{op})^2(1-\rho_{s,pk}^2\rho_{\varepsilon}^2)$. However, consumers are overly sensitive to the signal s_{it} when forming expectations about off-peak tastes. While the true conditional expectation is $E[\varepsilon_{it}^{op}|s] = \rho_{s,pk}\rho_{\varepsilon}\sigma_{\varepsilon}^{op}s_{it}$, consumers misperceive it to be $\frac{1}{\delta}\rho_{s,pk}\rho_{\varepsilon}\sigma_{\varepsilon}^{op}s_{it}$. An alternate assumption is that consumers also underestimate the off-peak unconditional standard deviation $\sigma_{\varepsilon}^{op}$ by δ. In this alternative, consumers are equally biased about the variance of peak and off-peak tastes but have correct conditional expectations for both peak and off-peak tastes. This alternative specification yields similar results.

(Note that this expression relies on our restriction $\rho_{s,op} = \rho_\varepsilon \rho_{s,pk}$.)

Note that, in the text, we focus on the marginal distribution of θ_{it}^{pk} and define both $\tilde{\mu}_{\theta it}^{pk} = \tilde{E}\left[\theta_{it}^{pk} \mid s_{it}\right]$ in equation (6) and $\tilde{\sigma}_{\theta t}^2 = \tilde{V}\left[\theta_{it}^{pk} \mid s_{it}\right]$ in equation (7), where we have factored out δ^2 given $\tilde{\sigma}_\varepsilon^{pk} = \delta \sigma_\varepsilon^{pk}$ and $\tilde{\sigma}_t = \delta \sigma_t$.

Learning from past usage At the end of billing period t, consumer i learns

$$z_{it} = \tilde{\theta}_{it}^{pk} - \rho_{s,pk} \sigma_\varepsilon^{pk} s_{it},$$

which she believes has distribution

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

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 (substituting $\tilde{\sigma}_1 = \delta \sigma_1$, $\sigma_1 = \sigma_\mu^{pk} \sqrt{1 - \rho_{\tilde{\mu},pk}^2}$, $\tilde{\sigma}_\varepsilon^{pk} = \delta \sigma_\varepsilon^{pk}$, and factoring out δ)

$$\tilde{\mu}_{i,t+1}^{pk} = \frac{\tilde{\mu}_{i1}^{pk} \left(1 - \rho_{\tilde{\mu},pk}^2\right)^{-1} (\sigma_\mu^{pk})^{-2} + t \bar{z}_{it} \left(1 - \rho_{s,pk}^2\right)^{-1} (\sigma_\varepsilon^{pk})^{-2}}{\left(1 - \rho_{\tilde{\mu},pk}^2\right)^{-1} (\sigma_\mu^{pk})^{-2} + t \left(1 - \rho_{s,pk}^2\right)^{-1} (\sigma_\varepsilon^{pk})^{-2}}, \tag{17}$$

$$\sigma_{t+1}^2 = \left(\left(1 - \rho_{\tilde{\mu},pk}^2\right)^{-1} (\sigma_\mu^{pk})^{-2} + t \left(1 - \rho_{s,pk}\right)^{-1} (\sigma_\varepsilon^{pk})^{-2}\right)^{-1}, \tag{18}$$

and

$$\tilde{\sigma}_{t+1} = \delta \sigma_{t+1}.$$

Note that $\tilde{\sigma}_{t+1}$ is always underestimated by a factor δ in each period, but that $\tilde{\mu}_{i,t+1}^{pk}$ is updated appropriately because $\tilde{\sigma}_\varepsilon^{pk}$ is underestimated by the same factor δ.

C Complete Model

The complete model differs from the illustrative version presented in the main text by accounting for piecewise-linear demand, inactive accounts, graduation, and free in-network calling.

C.1 Piecewise-linear Demand

We estimate $\beta = 2.7$. For any calling threshold above $1/\beta = \$0.37$, the value function defined in equation (2) leads to zero calling independent of θ because the maximum call value is $V_q(0, \theta) = 1/\beta$.

This prediction is not only unrealistic but also complicates estimation as it implies a $0.37 upper bound on v^* and hence an upper bound on s_{it} for all observations with positive usage. As a result, in the model we estimate, we adjust the value function to ensure that the first 1 percent of calling opportunities have an average value of at least $0.50. In particular, we assume that the inverse demand curve for calling minutes, $V_q(q_{it}^k, \theta_{it}^k)$, is:

$$V_q\left(q_{it}^k, \theta_{it}^k\right) = \max\left\{\left(\frac{1}{\beta_1} - \frac{1}{\beta_1} q_{it}^k/\theta_{it}^k\right), \left(1 - \frac{1}{\beta_2} q_{it}^k/\theta_{it}^k\right)\right\},$$

where $\beta_1 = \beta$ and $\beta_2 = 0.01$. (If we estimate β_2 the estimate does not move from the starting value $\beta_2 = 0.01$ and has a large standard error.) This inverse demand curve is illustrated in Figure 11. As before, demand is $q(v_{itj}^k, \theta_{it}^k) = \theta_{it}^k \hat{q}(v_{itj}^k)$. Now, however, $\hat{q}(v_{itj}^k)$ is

$$\hat{q}(v_{itj}^k) = \max\left\{\left(1 - \beta_1 v_{itj}^k\right), \beta_2\left(1 - v_{itj}^k\right)\right\}.$$

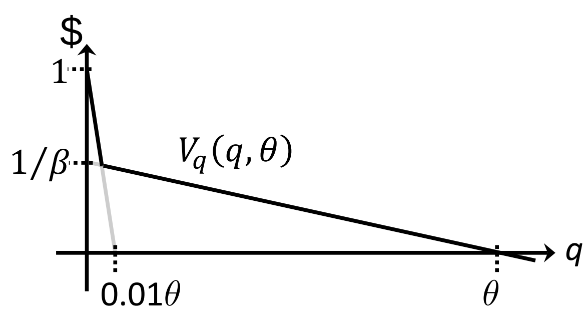

Figure 11: Piecewise-Linear Inverse Demand Curve

The corresponding value function is somewhat more cumbersome to express. It is

$$V\left(q_{it}^k, \theta_{it}^k\right) = \begin{cases} V^{II}\left(q_{it}^k, \theta_{it}^k\right), & q_{it}^k \leq q^*\left(\theta_{it}^k\right) \\ V^{I}\left(q_{it}^k, \theta_{it}^k\right) + \left(V^{II}\left(q^*, \theta_{it}^k\right) - V^{I}\left(q^*, \theta_{it}^k\right)\right), & q_{it}^k \geq q^*\left(\theta_{it}^k\right) \end{cases},$$

where $V^I\left(q_{it}^k, \theta_{it}^k\right) = q_{it}^k(\frac{1}{\beta_1} - \frac{1}{2\beta_1}\left(q_{it}^k/\theta_{it}^k\right))$, $V^{II}\left(q_{it}^k, \theta_{it}^k\right) = q_{it}^k(1 - \frac{1}{2\beta_2}\left(q_{it}^k/\theta_{it}^k\right))$, and $q^*\left(\theta_{it}^k\right) = \theta_{it}^k \frac{1-1/\beta_1}{1/\beta_2 - 1/\beta_1}$.

C.2 Inactive Accounts and Graduation

Consumers have the ability to put their accounts into inactive status, during which phones are disabled and fees are zero, and many do so over summer vacations. To account for these inactive periods we assume that no learning occurs during inactive periods and no taste shocks are observed. By letting time t index the number of active periods to date, formulas for beliefs based on Bayes rule in Appendix B.3 remain correct.

A substantial fraction of students terminate service at the end of the spring quarter. We assume that these terminations are exogenous because students were required to transfer service from the university to the firm upon graduation.

C.3 Free in-network calling

Calling plans distinguish between in-network and out-of-network calls as well as between peak and off-peak calls.[3] 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

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

[3] We matched area codes and exchanges of outgoing calls to carriers to determine in-network status using TelcoData.us data (TelcoData.us, 2005).

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

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

and, at the end of the month, realized usage in category k is $q^k_{it} = x^k_{it}\hat{q}(v^k_{itj})$. 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

$$U_{itj} = E\left[\sum_{k\in\{\text{pk-in,pk-out,op-in,op-out}\}} V\left(q^k(v^k_{itj}, x^k_{it}), x^k_{it}\right) - P_j\left(\mathbf{q}(\mathbf{v}^*_{itj}, \mathbf{x}_{it})\right) \mid \Im_{it}\right] + \eta_{itj}. \tag{19}$$

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

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

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: $\mathbf{v}^*_{itj} = (.11, 0, 0, 0)$. Second, plan 0 in fall 2003 or later when only in-network was free: $\mathbf{v}^*_{itj} = (.11, 0, .11, 0)$. Third, three-part tariffs with free in-network calling, such as plan 2 in January 2004: $\mathbf{v}^*_{itj} = (v^{pk,out}_{itj}, 0, 0, 0)$ and

$$v^{pk,out}_{itj} = p_j \Pr\left(x^{pk,out}_{it} \geq Q_j/\hat{q}(v^{pk,out}_{itj})\right) \frac{E\left[x^{pk,out}_{it} \mid x^{pk,out}_{it} \geq Q_j/\hat{q}(v^{pk,out}_{itj});\, \Im_{it}\right]}{E\left[x^{pk,out}_{it} \mid \Im_{it}\right]}. \tag{20}$$

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

$$v^{pk}_{itj} = p_j \Pr\left(\theta^{pk}_{it} \geq Q_j/\hat{q}(v^{pk}_{itj})\right) \frac{E\left[\theta^{pk}_{it} \mid \theta^{pk}_{it} \geq Q_j/\hat{q}(v^{pk}_{itj});\, \Im_{it}\right]}{E\left[\theta^{pk}_{it} \mid \Im_{it}\right]}. \tag{21}$$

As in the main text (Section IV), we break out calling demand for weekday outgoing calls to landlines immediately before and after 9pm to help identify the price sensitivity parameter. Now, however, such calls are a subset of out-of-network calling. Thus the shock $\mathbf{r}^{9pm}_{it} = (r^{pk,9}_{it}, r^{op,9}_{it}) \in [0,1]^2$ 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:

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

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:

$$E\left[r_{it}^{pk,9}\right] E\left[1 - r_{it}^{pk,in}\right] E\left[\theta_{it}^{pk}\right] = E\left[r_{it}^{op,9}\right] E\left[1 - r_{it}^{op,in}\right] E\left[\theta_{it}^{op}\right], \tag{22}$$

in place of equation (9). Equation (22) implicitly defines $\alpha_i^{op,9}$ as a function of $\alpha_i^{pk,9}$ and other parameters.

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

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

where α_i^k is unobserved heterogeneity and e_{it}^k is a mean-zero shock normally distributed with variance $(\sigma_e^k)^2$ independent across i, t, and k. For $k \in \{\text{pk-in,op-in,pk-9}\}$ 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 $\alpha_i^{k,in}$ and the distribution of $e_{it}^{k,in}$ for $k \in \{\text{pk,op}\}$ 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 $\delta_r r_{it}^{k,in}$. We incorporate this additional bias to help explain consumers choice of plan 1 over plan 0, which offered free in-network calling. Modeling in-network calling adds seven additional parameters to be estimated. These estimates are reported in Table 13. The parameters μ_α^{pk} and μ_α^{op} govern the average shares of $\boldsymbol{\theta}_{it}$ that can be apportioned to peak and off-peak in-network usage, respectively, while the next four govern the standard deviation of in-network usage shares between and within individuals. Since our estimate of δ_r is close to zero, consumers believe that almost all usage is out of network.

Table 13: In-Network Calling Parameter Estimates

Parameter	Description	Est.	Std. Err	Parameter	Description	Est.	Std. Err
μ_α^{pk}	$E[\alpha_i^{pk,in}]$	0.347	(0.01)	σ_e^{pk}	$SD[e_{it}^{pk,in}]$	0.183	(0.003)
μ_α^{op}	$E[\alpha_i^{op,in}]$	0.398	(0.009)	σ_e^{op}	$SD[e_{it}^{op,in}]$	0.168	(0.004)
σ_α^{pk}	$SD[\alpha_i^{pk,in}]$	0.188	(0.009)	δ_r	In-network bias	0.000	(0.004)
σ_α^{op}	$SD[\alpha_i^{op,in}]$	0.174	(0.01)				

C.4 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 $\delta_r E[r_{it}^{pk,in} \mid \alpha_i^{pk,in}]$. Thus initial plan-choice shares depend on the population distribution of $\delta_r E[r_{it}^{pk,in} \mid \alpha_i^{pk,in}]$. First consider a restricted model in which consumers are unbiased about in-network calling shares ($\delta_r = 1$). Then the population distribution of $\delta_r E[r_{it}^{pk,in} \mid \alpha_i^{pk,in}]$ is identified without knowing beliefs using data prior to fall 2003. During this period, all plans offered free nights-and-weekends so that $r_{it}^{op,in} = q_{it}^{op,in}/q_{it}^{op}$. 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 $r_{it}^{pk,in} = q_{it}^{pk,in}/q_{it}^{pk}$. For plan 0, peak calling-thresholds are 0 cents in-network and 11 cents out-of-network and hence

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

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

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 $r_{it}^{op,9}$ and $r_{it}^{pk,9}$ which are now computed as: $r_{it}^{op,9} = q_{it}^{op,9}/q_{it}^{op,out}$ and $r_{it}^{pk,9} = q_{it}^{pk,9}/q_{it}^{pk,out}$.

In an unreported robustness check, we find that a specification estimated with δ_r restricted to

1 significantly over-predicts the share of plan 0 after the plan stops offering free off-peak calling. Hence we allow consumers to underestimate their in-network calling share. To separately identify δ_r from overconfidence (δ) it is important to use both pre and post fall-2003 plan-choice-shares. Reducing δ_r or δ 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.

D Estimation Details

D.1 Model Fit

As shown in Table 14, our model does a good job of fitting plan choice shares. We also fit the monthly rate of switching well, which we predict to be 4.5 percent and is observed to be 3.7 percent. We note that our model is not flexible enough to capture the entire shape of the usage distribution. In Figure 12, we plot observed and predicted usage densities for both peak and off peak usage. The model tries to fit the shape of the usage distribution as closely as it can, but the censored normal specification we use produces a hump near zero that is not replicated in the data. A more flexible usage specification, such as a mixture of normals, might fit the observed data better.

Table 14: Plan Shares for Initial Choices (percent)

	October 2002				October 2003			
	Plan 0	Plan 1	Plan 2	Plan 3	Plan 0	Plan 1	Plan 2	Plan 3
Observed	67	10	21	2	18	72	8	1
Predicted	62	7	30	0.7	11	73	15	1

D.2 Simplified Likelihood Formulation

We begin this section by describing the structure of the likelihood function which arises from our model. To simplify the exposition, we initially make several simplifying assumptions. In particular, we ignore for the moment in-network and 8:00 pm to 10:00 pm usage shares, censoring of $\tilde{\boldsymbol{\theta}}_{it}$, and the fact that the set of firms considered by i (F_{it}) is unobserved. Additional details about the likelihood function, including censoring of $\tilde{\theta}^k_{it}$, a treatment of in-network calling, 8:00 pm to 10:00pm calling, and quitting, are in Appendix D.3. As discussed below, the likelihood function for our model does not have a closed form expression due to the presence of unobserved heterogeneity. We therefore

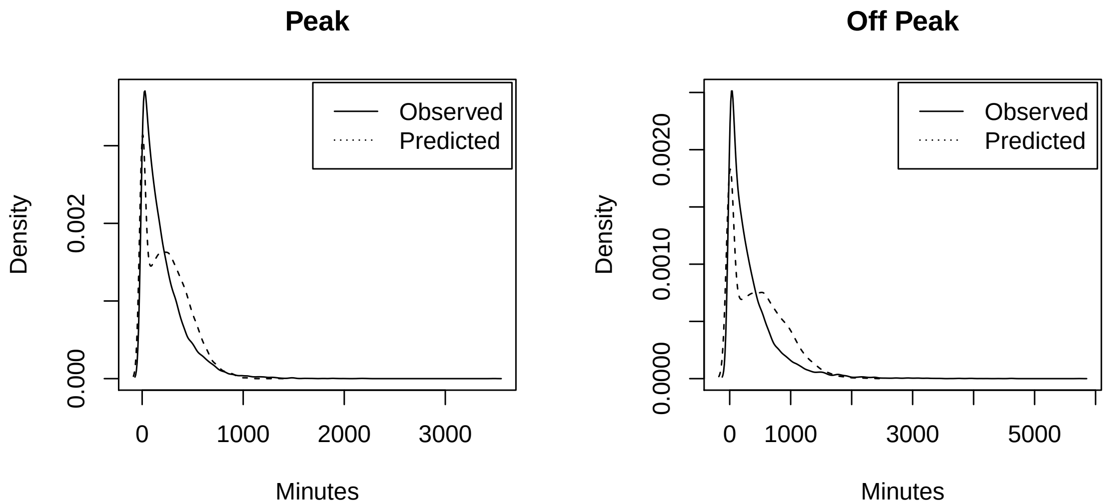

Figure 12: Observed and Predicted Usage Densities (Kernel Density Fits)

turn to Simulated Maximum Likelihood to approximate the likelihood function (Gourieroux and Monfort, 1993). Simulation details are given in Appendix D.4.

An observation in our model is a plan-choice and usage pair for a consumer at a given date: $\{j_{it}, \mathbf{q}_{it}\}$. For individual i at time period t, we observe a plan choice j_{it} as well as a vector of usage, $\boldsymbol{q}_{it}$, where $\boldsymbol{q}_{it} = \{q_{it}^{pk}, q_{it}^{op}\}$. We denote the sequence of observations for i up to time t by $\{\mathbf{j}_i^t, \mathbf{q}_i^t\}$, all observations for an individual i by $\{\mathbf{j}_i^T, \mathbf{q}_i^T\}$ and the full data set by $\{\mathbf{j}^T, \mathbf{q}^T\}$.

We construct the likelihood function from the distributional assumptions on our model's unobservables. To facilitate the exposition, we divide the unobservables into individual type $\boldsymbol{\omega}_i$, signals s_{it}, and taste shocks $\boldsymbol{\eta}_{it}$ and $\boldsymbol{\varepsilon}_{it}$. By individual type $\boldsymbol{\omega}_i$, we mean the vector of unobservables that are independent across individuals but constant across time: $\tilde{\mu}_{i1}^{pk}$, μ_i^{pk}, μ_i^{op}. The taste shocks are independent across individuals and time, and include the vector of firm-specific logit errors $(\boldsymbol{\eta}_{it})$ and the shocks governing total peak and off-peak usage $(\tilde{\boldsymbol{\theta}}_{it})$.

For each observation $\{j_{it}, \mathbf{q}_{it}\}$, we first evaluate the joint likelihood of observed plan choice and usage conditional on (1) consumer type ω_i, (2) past observations and signals $\{\mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}, \mathbf{s}_i^{t-1}\}$, and (3) the current signal s_{it}. This likelihood is denoted by $L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_i, \mathbf{s}_i^t, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right)$, where we suppress the dependence on Θ in our notation. The likelihood for an individual i is then constructed by taking the product of period t likelihoods and integrating over $\boldsymbol{\omega}_i$ and $\mathbf{s}_i^T$:

$$L_i(\mathbf{j}_i^T, \boldsymbol{q}_i^T) = \int_{\boldsymbol{\omega}_i} \int_{\mathbf{s}_i^T} \left(\prod_{t=1}^{T_i} L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_i, \mathbf{s}_i^t, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right) \right) f_s\left(\mathbf{s}_i^T\right) f_\omega\left(\boldsymbol{\omega}_i\right) d\mathbf{s}_i^T d\boldsymbol{\omega}_i. \tag{23}$$

As it has no closed form solution, we approximate the integral over $\boldsymbol{\omega}_i$ and $\mathbf{s}_i^T$ in equation (23) using Monte Carlo Simulation. For each individual, we take N_S draws on the random effects from $f_{\boldsymbol{\omega}}(\boldsymbol{\omega}_i)$ and $f_s\left(\boldsymbol{s}_i^T\right)$ and approximate the likelihood using

$$\hat{L}_i(\Theta) = \frac{1}{N_S}\sum_{b=1}^{N_S}\left(\prod_{t=1}^{T_i} L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_{ib}, \mathbf{s}_{ib}^t, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right)\right). \tag{24}$$

The simulated model log-likelihood is the sum of the logarithms of the individual simulated likelihoods:

$$\hat{LL}(\Theta) = \sum_{i=1}^{I}\log(\hat{L}_i(\Theta)). \tag{25}$$

We use $N_S = 400$ (randomly shuffled) Sobol simulation draws to calculate the simulated log-likelihood $\hat{LL}(\Theta)$ (see Appendix D.4 for details).

D.2.1 Conditional period t likelihood

We now construct the period t likelihood that enters equation (23). For convenience, we reexpress this likelihood conditional on the individual's peak type μ_i^{pk} and her information set $\Im_{it}$, which includes past choices $\mathbf{j}_i^{t-1}$ and $\mathbf{q}_i^{t-1}$, signals to date $\mathbf{s}_i^t$, and all elements of type $\boldsymbol{\omega}_i$ excluding μ_i^{pk}. Our construction follows naturally from the distributional assumptions on the taste shocks $\boldsymbol{\eta}_{it}$ and $\tilde{\boldsymbol{\theta}}_{it}$.

For a customer who quits, we only observe that $j_{it} \in J_{it}\backslash J_{it,university}$ and thus we sum: $\Pr(quit_{it} \mid \mu_i^{pk}, \Im_{it}) = \sum_{j_{it}\in J_{it}\backslash J_{it,university}} P_{it}(j_{it} \mid \Im_{it})$. Otherwise, conditional on $\{\mu_i^{pk}, \Im_{it}\}$, the likelihood of an observation $\{j_{it}, \mathbf{q}_{it}\}$ is the product of the plan choice probability and the likelihood of observed usage conditional on plan choice. In particular, conditional on s_{it}, we will infer $\tilde{\boldsymbol{\theta}}_{it}$ from observed $\boldsymbol{q}_{it}$ and write the likelihood as:

$$L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_i, \mathbf{s}_i^t, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right) = L_{it}(j_{it}, \boldsymbol{q}_{it} \mid \mu_i^{pk}, \Im_{it}) = \Pr(j_{it} \mid \Im_{it}) f_{\boldsymbol{q}}(\boldsymbol{q}_{it}|j_{it}, \mu_i^{pk}, \Im_{it}). \tag{26}$$

Usage likelihood: First, we consider the likelihood of the observed usage $\mathbf{q}_{it}$. We know that for $k \in \{pk, op\}$: $q_{it}^k = \theta_{it}^k \hat{q}(v_{it}^k(j_{it}, \Im_{it}))$. Therefore, conditional on $v_{it}^k(j_{it}, \Im_{it})$, we can infer θ_{it}^k from usage: $\theta_{it}^k(q_{it}^k, v_{it}^k(j_{it}, \Im_{it})) = q_{it}^k / \hat{q}(v_{it}^k(j_{it}, \Im_{it}))$. Then, the joint normal assumption on $\boldsymbol{\varepsilon}_{it}$ and s_{it} leads directly to the distribution of $\tilde{\boldsymbol{\theta}}_{it}$ conditional on s_{it}: $\tilde{\boldsymbol{\theta}}_{it} \sim N\left(\mu_{it}^k + E\left[\boldsymbol{\varepsilon}_{it} \mid s_{it}\right], V\left[\boldsymbol{\varepsilon}_{it} \mid s_{it}\right]\right)$ (See Appendix B.3). Assuming no censoring (so that $\boldsymbol{\theta}_{it} = \tilde{\boldsymbol{\theta}}_{it}$) this yields the usage likelihood:

$$f_{\boldsymbol{q}}(\boldsymbol{q}_{it}|j_{it}, \mu_i^{pk}, \Im_{it}) = f_{\tilde{\boldsymbol{\theta}}|s}(\boldsymbol{\theta}_{it}(\mathbf{q}_{it}, \mathbf{v}_{it}^*(j_{it}, \Im_{it})) \mid \mu_i, s_{it}) \left|\frac{\partial \boldsymbol{\theta}_{it}(\mathbf{q}_{it}, \mathbf{v}_{it}^*(j_{it}, \Im_{it})}{\partial \boldsymbol{q}_{it}}\right|, \tag{27}$$

where the term $|\partial\boldsymbol{\theta}_{it}(\mathbf{q}_{it}, \mathbf{v}^*_{it}(j_{it}, \Im_{it})/\partial\boldsymbol{q}_{it}|$ is the Jacobian determinant of the transformation between $\boldsymbol{\theta}_{it}$ and $\boldsymbol{q}_{it}$ (conditional on $\mathbf{v}^*_{it}(j_{it}, \Im_{it})$). The usage likelihood must be adjusted whenever an element of $\tilde{\boldsymbol{\theta}}_{it}$ is censored at zero, and in the complete model there are additional terms to incorporate data on in-network calling and 8-10pm calling (Appendix D.3).

Plan Choice Probability: Next we consider plan choice conditional on $\Im_{it}$ and an active choice, which is denoted by conditioning on "C". Let F_{it} denote the set of firms considered by the consumer, $J_{it}(f)$ denote the set of plans available from firm $f \in F_{it}$, and $f(j_{it})$ denote the firm which offers plan j_{it}. (Note that $J_{it}(f)$ depends on $j_{i,t-1}$ because consumers can always keep their existing plan.) Then the probability of choosing plan j_{it} equals the probability of choosing firm $f(j_{it})$ multiplied by the probability of choosing plan j_{it} conditional on choosing firm $f(j_{it})$:

$$\Pr(j_{it} \mid C; \Im_{it}, F_{it}) = \Pr(f(j_{it}) \mid C; \Im_{it}, F_{it}) \Pr(j_{it} \mid C; \Im_{it},\ f(j_{it}))$$

Conditional on an active choice, our assumption of firm-specific logit errors gives rise to the following firm choice probability:

$$\Pr(f \mid C; \Im_{it}, F_{it}) = \frac{\exp(\max_{k \in J_{it}(f)} \{U_{itk}(\Im_{it})\})}{\sum_{g \in F_{it}} \exp(\max_{k \in J_{it}(g)} \{U_{itk}(\Im_{it})\})}.$$

Conditional on an active choice of firm $f(j_{it})$, the probability of choosing plan j_{it} is simply equal to an indicator function specifying whether or not plan j_{it} offered the highest expected utility to consumer i of any plan offered by firm $f(j_{it})$:

$$\Pr(j_{it} \mid C; \Im_{it},\ f(j_{it})) = \mathbf{1}\left\{U_{itj}(\Im_{it}) = \max_{k \in J_{it}(f(j_{it}))} \{U_{itk}(\Im_{it})\}\right\}. \tag{28}$$

(Typically we characterize $\Pr(j_{it} \mid C; \Im_{it},\ f(j_{it}))$ by numerically computing bounds $\underline{s}_{it}$ and $\bar{s}_{it}$ such that it is equal to 1 if an only if $s_{it} \in [\underline{s}_{it}, \bar{s}_{it}]$.)

In period 1, every consumer makes an active choice from the set of university plans, so $\Pr(j_{i1} \mid \Im_{i1}) = \Pr(j_{i1} \mid C; \Im_{i1}, f(j_{i1}))$. In every later period, however, consumers keep their existing plan $j_{i,t-1}$ with probability $(1 - P_C)$ and otherwise choose between the university, the outside option, and one of three outside firms. Thus, unconditional on an active choice, the probabilities that an existing customer switches to plan $j \neq j_{i,t-1}$ in period t (where j could be the outside good) or keeps the existing plan $j = j_{i,t-1}$ are $P_C \Pr(j_{it} \mid C; \Im_{it}, F_{it})$ and $P_C \Pr(j_{it} \mid C; \Im_{it}, F_{it}) + (1 - P_C)$ respectively:

$$\Pr(j_{it} \mid \Im_{it}, F_{it}) = \begin{cases} P_C \Pr(j_{it} \mid C; \Im_{it}, F_{it}) & \text{if } j \neq j_{i,t-1} \\ P_C \Pr(j_{it} \mid C; \Im_{it}, F_{it}) + (1 - P_C) & \text{if } j = j_{i,t-1} \end{cases}. \tag{29}$$

Finally, the set of firms considered by consumers is unobserved. Each existing consumers considers university plans, the outside option, and one of the three outside firms offering local plans (AT&T, Cingular, or Verizon). Depending on the identity of the outside firm, there are three possible consideration sets: $F_{it} \in \left\{F_{it}^1, F_{it}^2, F_{it}^3\right\}$, each of which is equally likely. Thus the final plan choice probability is:

$$\Pr\left(j_{it} \mid \Im_{it}\right) = \frac{1}{3}\sum_{k=1}^{3} \Pr\left(j_{it} \mid \Im_{it}, F_{it}^k\right). \tag{30}$$

D.3 Complete Likelihood Formulation

In this section we revise the likelihood function described in Appendix D.2 to fully account for (1) censoring of taste shocks $\boldsymbol{\theta}_{it}$, (2) free in-network calling and in-network calling data, and (3) near 9pm calling data.

There are three changes that affect the individual i likelihood, for which equation (23) gives the simplified version. First, the type vector $\boldsymbol{\omega}_i$ is expanded to include parameters governing i's distribution of near 9pm and in-network usage: $\boldsymbol{\omega}_i = (\tilde{\mu}_{i1}^{pk}, \mu_i^{pk}, \mu_i^{op}, \alpha_i^{pk,9}, \alpha_i^{pk,in}, \alpha_i^{op,in})$. Second, the usage vector $\mathbf{q}_{it}$ is expanded to include in-network and near 9pm calling data:

$$\boldsymbol{q}_{it} = \{q_{it}^{pk,in}, q_{it}^{pk,out}, q_{it}^{pk,9}, q_{it}^{op,in}, q_{it}^{op,out}, q_{it}^{op,9}\}.$$

Third, censoring of the latent taste shocks means that there is an additional vector of unobservable variables. Let $\tilde{\boldsymbol{\theta}}_i^-$ be the vector of all latent taste shocks for individual i that are negative and hence unobserved (each corresponding to an observed $\theta_{it}^k = 0$). Let $\tilde{\boldsymbol{\theta}}_i^{-,t}$ be the subset of $\tilde{\boldsymbol{\theta}}_i^-$ realized at or before time t.

As before, we will construct the joint likelihood $\{j_{it}, \mathbf{q}_{it}\}$ by conditioning on $\{\mu_i^{pk}, \Im_{it}\}$. Without censoring, this was equivalent to conditioning on $\left\{\boldsymbol{\omega}_i, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}, \mathbf{s}_i^t\right\}$. Now, however, it requires also conditioning on $\tilde{\boldsymbol{\theta}}_i^{-,t-1}$ because consumer i's information set $\Im_{it}$ includes all elements of $\tilde{\boldsymbol{\theta}}_i^-$ realized before time t but these cannot be inferred from $\mathbf{q}_i^{t-1}$. Thus we must now integrate out over this additional unobserved heterogeneity:

$$\begin{aligned} L_i(\mathbf{j}_i^T, \boldsymbol{q}_i^T) \;\; = \;\; & \int_{\boldsymbol{\omega}_i}\int_{\mathbf{s}_i^T}\int_{\tilde{\boldsymbol{\theta}}_i^-} \left(\prod_{t=1}^{T_i} L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_i, \mathbf{s}_i^t, \tilde{\boldsymbol{\theta}}_i^{-,t-1}, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right)\right) \\ & f_{\tilde{\theta}}\left(\tilde{\boldsymbol{\theta}}_i^- | \mathbf{s}_i^T, \boldsymbol{\omega}_i, \tilde{\boldsymbol{\theta}}_i^- \leq 0\right) f_s\left(\mathbf{s}_i^T\right) f_\omega\left(\boldsymbol{\omega}_i\right) d\tilde{\boldsymbol{\theta}}_i^- d\mathbf{s}_i^T d\boldsymbol{\omega}_i. \end{aligned} \tag{31}$$

The expression for the likelihood $L_{it}(j_{it}, \boldsymbol{q}_{it} \mid \mu_i^{pk}, \Im_{it})$ given by equation (26) must also be revised. In particular, we replace the density of $\mathbf{q}_{it}$, f_q, with the likelihood of $\mathbf{q}_{it}$, l^q:

$$L_{it}(j_{it}, \boldsymbol{q}_{it} \mid \mu_i^{pk}, \Im_{it}) = P_{it}(j_{it} \mid \Im_{it}) l^q(\boldsymbol{q}_{it} | j_{it}, \mu_i^{pk}, \Im_{it}). \tag{32}$$

The remainder of this section is devoted to fully specifying the likelihood of observed usage, $l^q(\boldsymbol{q}_{it}|j_{it}, \mu_i^{pk}, \Im_{it})$.

The usage vector $\boldsymbol{q}_{it}$ is a function of the random variables $\boldsymbol{\theta}_{it}$ and $\mathbf{r}_{it} = \{r_{it}^{pk,in}, r_{it}^{op,in}, r_{it}^{pk,9}, r_{it}^{op,9}\}$. To compute the likelihood $l^q(\boldsymbol{q}_{it} \mid j_{it}, \mu_i^{pk}, \Im_{it})$, we compute the likelihood of $\boldsymbol{\theta}_{it}$, $l^\theta(\boldsymbol{\theta}_{it} \mid j_{it}, \mu_i^{pk}, \Im_{it})$ and the likelihood of $\mathbf{r}_{it}$, $l^r(\mathbf{r}_{it} \mid j_{it}, \Im_{it})$, put them together and then do the change of variables from these variables to $\boldsymbol{q}_{it}$:

$$l^q(\boldsymbol{q}_{it} \mid j_{it}, \mu_i^{pk}, \Im_{it}) = l^\theta(\boldsymbol{\theta}_{it}|j_{it}, \mu_i^{pk}, \Im_{it}) l^r(\mathbf{r}_{it}|j_{it}, \Im_{it}) \left| \det \left(\frac{\partial (\boldsymbol{\theta}_{it}, \mathbf{r}_{it})}{\partial \boldsymbol{q}_{it}} \right) \right|.$$

Below we (1) outline the construction of $l^\theta(\boldsymbol{\theta}_{it}|j_{it}, \mu_i^{pk}, \Im_{it})$ which accounts for censoring, (2) outline the construction of $l^r(\mathbf{r}_{it}|j_{it}, \Im_{it})$ by incorporating (a) in-network calling data, and (b) near 9-pm calling data, and (3) finish by outlining $|\det(\partial (\boldsymbol{\theta}_{it}, \mathbf{r}_{it}) / \partial \boldsymbol{q}_{it})|$ for the change of variables.

D.3.1 Censoring

Conditional on $\boldsymbol{\mu}_i$ and s_{it}, the taste shock $\tilde{\boldsymbol{\theta}}_{it}$ follows a bivariate normal distribution:

$$\tilde{\boldsymbol{\theta}}_{it} \sim N\left(\mu_{it}^k + E[\boldsymbol{\varepsilon}_{it} \mid s_{it}], V[\boldsymbol{\varepsilon}_{it} \mid s_{it}]\right)$$

where $E[\boldsymbol{\varepsilon}_{it} \mid s_{it}]$ and $V[\boldsymbol{\varepsilon}_{it} \mid s_{it}]$ follow from Bayes rule in Appendix B.3 equations (15)-(16). We denote the joint density by $f_{\tilde{\theta}|s}(\tilde{\boldsymbol{\theta}}_{it}|\boldsymbol{\mu}_i, s_{it})$, and the marginal density of θ_{it}^k for $k \in \{pk, op\}$ by $f_{\tilde{\theta}|s}^k(\tilde{\theta}_{it}^k|\mu_i^k, s_{it})$.

As described in Section V, θ_{it}^k can be inferred from observed usage. If $\theta_{it}^k > 0$, then this gives the value of the latent variable: $\tilde{\theta}_{it}^k = \theta_{it}^k$. If $\theta_{it}^k = 0$, however, we can only infer that $\tilde{\theta}_{it}^k \leq 0$. When $\boldsymbol{\theta}_{it}$ is not censored, its likelihood is simply $f_{\tilde{\boldsymbol{\theta}}|s}(\boldsymbol{\theta}_{it}(\mathbf{q}_{it}, \mathbf{v}_{it}^*(j_{it}, \Im_{it}))$. Otherwise, the likelihood of $\boldsymbol{\theta}_{it}(\mathbf{q}_{it}, \mathbf{v}_{it}^*(j_{it}, \Im_{it})$ must be adjusted by substituting the probability that $\tilde{\theta}_{it}^k$ is censored for $f_{\boldsymbol{\theta}|\boldsymbol{s}}$:

$$l^\theta(\boldsymbol{\theta}_{it} \mid j_{it}, \mu_i^{pk}, \Im_{it}) = \begin{cases} f_{\tilde{\theta}|s}(\boldsymbol{\theta}_{it}|\boldsymbol{\mu}_i, s_{it}) & \text{if } \theta_{it}^{pk} > 0, \theta_{it}^{op} > 0 \\ \Pr(\tilde{\theta}_{it}^{pk} \leq 0|\mu_i^{pk}, s_{it}, \theta_{it}^{op}) f_{\tilde{\theta}|s}^k(\theta_{it}^{op}|\mu_i^{op}, s_{it}) & \text{if } \theta_{it}^{pk} = 0, \theta_{it}^{op} > 0 \\ \Pr(\tilde{\theta}_{it}^{op} \leq 0|\mu_i^{op}, s_{it}, \theta_{it}^{pk}) f_{\tilde{\theta}|s}^k(\theta_{it}^{pk}|\mu_i^{pk}, s_{it}) & \text{if } \theta_{it}^{pk} > 0, \theta_{it}^{op} = 0 \\ \Pr(\tilde{\theta}_{it}^{pk} \leq 0, \tilde{\theta}_{it}^{op} \leq 0|\boldsymbol{\mu}_i, s_{it}) & \text{if } \theta_{it}^{pk} = 0, \theta_{it}^{op} = 0 \end{cases}, \tag{33}$$

where $\boldsymbol{\theta}_{it}$'s dependence on $\mathbf{q}_{it}$ and $\mathbf{v}^*_{it}(j_{it}, \Im_{it})$ is suppressed from the notation. We then use the change of variables formula to arrive at the likelihood of q_{it} (rather than $\boldsymbol{\theta}_{it}$) We leave this step, however, until after considering in-network and near-9pm calling.

D.3.2 In-network and Near-9pm Calling

By including data on whether calls are in or out of network and whether calls are within 60 minutes of 9pm, the usage vector becomes $\boldsymbol{q}_{it} = \{q^{pk,in}_{it}, q^{pk,out}_{it}, q^{pk,9}_{it}, q^{op,in}_{it}, q^{op,out}_{it}, q^{op,9}_{it}\}$, where "*in*", "*out*", and "9" signify in-network, out-of-network, and near-9pm calling respectively. As before, $q^{pk}_{it} = q^{pk,in}_{it} + q^{pk,out}_{it}$ and $q^{op}_{it} = q^{op,in}_{it} + q^{op,out}_{it}$ denote total peak and total off-peak calling respectively. In this section we discuss three issues related to handling in-network calling: (1) a data limitation, (2) the inference of $\boldsymbol{\theta}_{it}$, and (3) an additional term in the likelihood.

(1) Data Limitation: We always observe total peak and off-peak calling because we observe the time and date of all calls. An important data limitation, however, is that call logs only directly identify outgoing calls as in-network or out-of-network. This information provides lower and upper bounds on in-network calling: $\underline{q}^{k,in}_{it}$ and $\bar{q}^{k,in}_{it}$ for $k \in \{pk, op\}$. 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. Analogous bounds on out-of-network usage are $\underline{q}^{k,out}_{it} = q^k_{it} - \bar{q}^{k,in}_{it}$ and $\bar{q}^{k,out}_{it} = q^k_{it} - \underline{q}^{k,in}_{it}$ for $k \in \{pk, op\}$. 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 $q^{pk,in}_{it}$ and $q^{op,in}_{it}$ exactly.

(2) **Inferring $\boldsymbol{\theta}_{it}$**: The fact that in-network and out-of-network calling may be priced differently complicates our inference of $\boldsymbol{\theta}_{it}$ from usage data. For $k \in \{pk, op\}$, θ^k_{it} is calculated by equation (34) if category k calls are not priced differentially by network status or by equation (35) if category k calls are priced differentially by network status:

$$\theta^k_{it} = q^k_{it}/\hat{q}(v^k_{it}), \tag{34}$$

$$\theta^k_{it} = q^{k,in}_{it}/\hat{q}(v^{k,in}_{it}) + q^{k,out}_{it}/\hat{q}(v^{k,out}_{it}). \tag{35}$$

There is always sufficient information to infer $\boldsymbol{\theta}_{it}$ from usage conditional on the threshold vector $\mathbf{v}^*_{it}(j_{it}, \Im_{it})$ because $q^{k,in}_{it}$ and $q^{k,out}_{it}$ are observed precisely when equation (35) applies.

(3) Likelihood Function: Next we turn to the likelihood of in-network and near 9pm calling shares.[4] For $k \in \{pk, op\}$, if in-network usage is observed exactly then we can calculate the exact share of category k calling opportunities that are in-network as

$$r_{it}^{k,in} = \frac{q_{it}^{k,in}/\hat{q}(v_{it}^{k,in})}{q_{it}^{k,in}/\hat{q}(v_{it}^{k,in}) + q_{it}^{k,out}/\hat{q}(v_{it}^{k,out})},$$

and the exact share of category k calling opportunities that are near-9pm as

$$r_{it}^{k,9} = q_{it}^{k,9}/q_{it}^{k,out}.$$

For $k \in$ {pk-in,pk-9,op-in,op-9}, as r_{it}^{k} follows a censored-normal distribution, where the underlying normal distribution is defined by $\alpha_i^k + e_{it}^k$, we can write the likelihood of r_{it}^k as:

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

For $k \in \{pk, op\}$, if we only observe bounds on $q_{it}^{k,in}$ and $q_{it}^{k,out}$ then we can only calculate bounds for $r_{it}^{k,in}$ and $r_{it}^{k,9}$:

$$\begin{aligned} r_{it}^{k,in} &\in \left[\frac{\underline{q}_{it}^{k,in}}{\theta_{it}^k \hat{q}(v_{it}^{k,in})}, \frac{\bar{q}_{it}^{k,in}}{\theta_{it}^k \hat{q}(v_{it}^{k,in})}\right] = \left[\underline{r}_{it}^{k,in}, \bar{r}_{it}^{k,in}\right], \\ r_{it}^{k,9} &\in \left[\frac{q_{it}^{k,9}}{\bar{q}_{it}^{k,out}}, \frac{q_{it}^{k,9}}{\underline{q}_{it}^{k,out}}\right] = \left[\underline{r}_{it}^{k,9}, \bar{r}_{it}^{k,9}\right]. \end{aligned}$$

Note that $q_{it}^{k,9} \le \underline{q}_{it}^{k,out}$, so $q_{it}^{k,9} > 0$ implies the bounds on $r_{it}^{k,9}$ are within $[0,1]$. For $k \in \{pk, op\}$, if the upper bound on the category k in-network usage share is below one ($\bar{r}_{it}^{k,in} < 1$) then the likelihood for the bounds on $r_{it}^{k,in}$ and $r_{it}^{k,9}$ are as follows.

$$l^{k,in}(\underline{r}_{it}^{k,in},\ \bar{r}_{it}^{k,in} \mid \alpha_i^k) = \begin{cases} f^k(r_{it}^{k,in}|\alpha_i^{k,in}) & \text{if } \underline{r}_{it}^{k,in} = \bar{r}_{it}^{k,in} = r_{it}^{k,in} \\ \Phi\left(\left(\bar{r}_{it}^{k,in} - \alpha_i^{k,in}\right)/\sigma_e^{k,in}\right) - \Phi\left(\left(\underline{r}_{it}^{k,in} - \alpha_i^{k,in}\right)/\sigma_e^{k,in}\right) & \text{if } 0 < \underline{r}_{it}^{k,in} < \bar{r}_{it}^{k,in} < 1 \\ \Phi\left(\left(\bar{r}_{it}^{k,in} - \alpha_i^{k,in}\right)/\sigma_e^{k,in}\right) & \text{if } 0 = \underline{r}_{it}^{k,in} < \bar{r}_{it}^{k,in} < 1 \end{cases}$$

[4] Note that, for $k \in \{pk, op\}$, when $q_{it}^k = 0$ we have no information about the share of in-network usage $r_{it}^{k,in}$ and when $q_{it}^{k,out} = 0$ and we have no information about the share of near-9pm $r_{it}^{k,9}$.

$$l^{k,9}(\underline{r}_{it}^{k,9},\ \bar{r}_{it}^{k,9} \mid \alpha_i^{k,9}) = \begin{cases} f^{k,9}(r_{it}^{k,9}|\alpha_i^{k,9}) & \text{if } \underline{r}_{it}^{k,9} = \bar{r}_{it}^{k,9} = r_{it}^{k,9} \\ \Phi\left(\left(\bar{r}_{it}^{k,9} - \alpha_i^{k,9}\right)/\sigma_e^{k,9}\right) - \Phi\left(\left(\underline{r}_{it}^{k,9} - \alpha_i^{k,9}\right)/\sigma_e^{k,9}\right) & \text{if } 0 < \underline{r}_{it}^{k,9} < \bar{r}_{it}^{k,9} < 1 \\ \Phi\left(\left(\bar{r}_{it}^{k,9} - \alpha_i^{k,9}\right)/\sigma_e^{k,9}\right) & \text{if } 0 = \underline{r}_{it}^{k,9} < \bar{r}_{it}^{k,9} < 1 \\ 1 - \Phi\left(\left(\underline{r}_{it}^{k,9} - \alpha_i^{k,9}\right)/\sigma_e^{k,9}\right) & \text{if } 0 \le \underline{r}_{it}^{k,9} < \bar{r}_{it}^{k,9} = 1 \end{cases}$$

Moreover, in the case $\bar{r}_{it}^{k,in} < 1$, the likelihoods of $r_{it}^{k,in}$ and $r_{it}^{k,9}$ are independent and we can write the joint likelihood as the product of $l^{k,in}(\underline{r}_{it}^{k},\ \bar{r}_{it}^{k} \mid \alpha_i^{k})$ and $l^{k,9}(\underline{r}_{it}^{k,9},\ \bar{r}_{it}^{k,9} \mid \alpha_i^{k,9})$.

When $\bar{r}_{it}^{k,in} = 1$, constructing the likelihood of $r_{it}^{k,in}$ and $r_{it}^{k,9}$ is more complicated. The complication 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 (if in-network calling is free) 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 near 9pm could be zero for two reasons: $r_{it}^{k,in} = 1$, or $r_{it}^{k,9} = 0$. If the upper bound on $r_{it}^{k,in}$ binds, then $r_{it}^{k,9}$ could take any value. However, if the upper bound on $r_{it}^{k,in}$ does not bind, then $r_{it}^{k,9}$ must be zero. Following this logic, the joint likelihood of $\underline{r}_{it}^{k,in} \le r_{it}^{k,in} \le 1$ and $r_{it}^{k,9} \in [0,1]$ is

$$l^{r,k}\left(\underline{r}_{it}^{k,in}, \bar{r}_{it}^{k,in} = 1, r_{it}^{k,9} \in [0,1] \mid \alpha_i^{k,in}, \alpha_i^{k,9}\right)$$

$$= \begin{cases} 1 - \Phi((1-\alpha_i^{k,in})/\sigma_e^{k,in}) + \Phi(-\alpha_i^{k,9}/\sigma_e^{k,9}) \begin{pmatrix} \Phi((1-\alpha_i^{k,in})/\sigma_e^{k,in}) \\ -\Phi((\underline{r}_{it}^{k,in} - \alpha_i^{k,in})/\sigma_e^{k,in}) \end{pmatrix} & \text{if } \underline{r}_{it}^{k,in} > 0 \\ 1 - \Phi((1-\alpha_i^{k,in})/\sigma_e^{k,in}) + \Phi(-\alpha_i^{k,9}/\sigma_e^{k,9})\Phi((1-\alpha_i^{k,in})/\sigma_e^{k,in}) & \text{if } \underline{r}_{it}^{k,in} = 0. \end{cases}$$

Thus the joint likelihood of in-network and near-9pm calling opportunity shares for category $k \in \{pk, op\}$ is

$$l^{r,k}(\underline{r}_{it}^{k,in}, \bar{r}_{it}^{k,in}, \underline{r}_{it}^{k,9}, \bar{r}_{it}^{k,9} \mid \alpha_i^{k,in}, \alpha_i^{k,9}) = \begin{cases} l^{r,k}(\underline{r}_{it}^{k,in}, \bar{r}_{it}^{k,in} \mid \alpha_i^{k,in}) \cdot l^{k,9}(\underline{r}_{it}^{k,9}, \bar{r}_{it}^{k,9} \mid \alpha_i^{k,9}) & \text{if } \bar{r}_{it}^{k,in} < 1 \\ l^{r,k}(\underline{r}_{it}^{k,in}, \bar{r}_{it}^{k,in} = 1, r_{it}^{k,9} \in [0,1] \mid \alpha_i^{k,in}, \alpha_i^{k,9}) & \text{if } \bar{r}_{it}^{k,in} = 1. \end{cases}$$

Finally,

$$l^r(\mathbf{r}_{it}|j_{it}, \Im_{it}) = \Pi_{k\in\{pk,op\}} l^{r,k}(\underline{r}_{it}^{k,in}, \bar{r}_{it}^{k,in}, \underline{r}_{it}^{k,9}, \bar{r}_{it}^{k,9} \mid \alpha_i^{k,in}, \alpha_i^{k,9}).$$

In the following section we perform the change of variables transformation to write the likelihood as a function of $\mathbf{q}_{it}$ rather than $\mathbf{r}_{it}$.

D.3.3 Change of Variables

The joint likelihood of $\boldsymbol{\theta}_{it}$ and $\mathbf{r}_{it}$ will be the product of the l^{θ} and l^{r}. This is not the likelihood of the observed usage vector $\boldsymbol{q}_{it}$, however, because $\boldsymbol{q}_{it}$ is a function of $\tilde{\boldsymbol{\theta}}_{it}$ and $\mathbf{r}_{it}$ 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}, $r^{k,in}$, $r^{k,9}$ for $k \in \{pk, op\}$ is

$$\theta^{k} = \frac{q^{k,out}}{\hat{q}(v^{k,out})} + \frac{q^{k,in}}{\hat{q}(v^{k,in})};\ r^{k,in} = \frac{q^{k,out}}{q^{k,out} + q^{k,in}\frac{\hat{q}(v^{k,out})}{\hat{q}(v^{k,in})}};\ r^{k,9} = \frac{q^{k,9}}{q^{k,out}}.$$

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, 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 $\boldsymbol{y} = (q^{pk,out}, q^{pk,in}, q^{pk,9}, q^{op,out}, q^{op,in}, q^{op,9})$ and $\boldsymbol{x} = (\theta^{pk}, r^{pk,in}, r^{pk,9}, \theta^{op}, r^{op,in}, r^{op,9})$,

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

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

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

and

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

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

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

The determinant of this Jacobian will be

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

D.4 Simulation Details

In this section we describe in detail the procedure we follow for approximating integrals in the likelihood using Monte Carlo Simulation.

D.4.1 Simulation Draws

It is well-known that the value of Θ which maximizes the simulated log-likelihood $\hat{LL}$ is inconsistent for a fixed number of simulation draws N_S due to the logarithmic transformation in equation (25). However, it is consistent if $N_S \to \infty$ as $I \to \infty$, as discussed in Hajivassiliou and Ruud (1994).

We chose $N_S = 400$; to arrive at this value we conducted some simple artificial data experiments where we simulated our model and attempted to recover the parameters, finding that 400 draws was sufficient to recover the true parameter draws to roughly 5 percent accuracy. We found that 400 draws per consumer was the maximum number of draws we could feasibly include in our estimation. We re-estimated our model with only 300 draws, and found that our parameter estimates were very similar to those obtained with 400 draws, suggesting that increasing the number of draws beyond 400 would not significantly improve the parameter estimates.

We also found in our experiments that we were able to reduce simulation bias significantly by using a deterministic Sobol sequence generator to create the random draws, rather than canonical random number generators. (We randomly shuffle the Sobol draws, independently for each element). Goettler and Shachar (2001) describe some of the advantages of this technique in detail. We use the algorithm provided in the R package randtoolbox to create the draws (Dutang and Savicky, 2010).

D.4.2 Approximate Likelihood Function

As it has no closed form solution, we approximate the integral over $\boldsymbol{\omega}_i$, $\mathbf{s}_i^T$, and $\tilde{\boldsymbol{\theta}}_i^-$ in equation (31) numerically using importance sampling and Monte Carlo Simulation. A natural first approach would be to take N_S draws on the random effects from $f_{\tilde{\theta}}\left(\tilde{\boldsymbol{\theta}}_i^- | \mathbf{s}_i^T, \boldsymbol{\omega}_i, \tilde{\boldsymbol{\theta}}_i^- \leq 0\right)$, $f_{\boldsymbol{\omega}}(\boldsymbol{\omega}_i)$, and $f_s\left(\boldsymbol{s}_i^T\right)$ and approximate individual i's likelihood using

$$\hat{L}_i(\Theta) = \frac{1}{N_S}\sum_{b=1}^{N_S}\left(\prod_{t=1}^{T_i} L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_{ib}, \mathbf{s}_{ib}^t, \tilde{\boldsymbol{\theta}}_{ib}^{-,t-1}, F_{itb}, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right)\right),$$

where b indexes the simulation draw. Unfortunately, this approach would lead the approximate likelihood function to be discontinuous. (This fact follows from equation (28) because $\Pr(j_{it} \mid C; \Im_{it}$, $f\left(j_{it}\right))$ is discontinuous in s_{it}.)

Our alternate approach is to compute the bounds $\underline{s}_{it}$ and $\bar{s}_{it}$ such that $\Pr(j_{it} \mid C; \Im_{it}, f\left(j_{it}\right)) = 1$ if an only if $s_{it} \in [\underline{s}_{it}, \bar{s}_{it}]$. (We discuss computation of the bounds in detail in the following section.) We then take N_S draws on s_{it} from a truncated normal within these bounds. Drawing s_{it} from the truncated density $f_s(s)/\Pr\left(s_{it} \in [\underline{s}_{it}, \bar{s}_{it}]\right)$ necessitates multiplying the likelihood expression L_{it} by $\Pr\left(s_{it} \in [\underline{s}_{it}, \bar{s}_{it}]\right)$ to cancel out the denominator when one takes the expectation of the approximation. Thus our approximate likelihood function is:

$$\hat{L}_i(\Theta) = \frac{1}{N_S}\sum_{b=1}^{N_S}\left(\prod_{t=1}^{T_i} L_{it}\left(j_{it}, \boldsymbol{q}_{it} \mid \boldsymbol{\omega}_{ib}, \mathbf{s}_{ib}^t, \tilde{\boldsymbol{\theta}}_{ib}^{-,t-1}, F_{itb}, \mathbf{j}_i^{t-1}, \mathbf{q}_i^{t-1}\right) \Pr\left(s_{it} \in [\underline{s}_{it}, \bar{s}_{it}]\right)\right), \qquad (36)$$

which is continuous. Given the bounds $[\underline{s}_{it}, \bar{s}_{it}]$, drawing s_{it} from a truncated normal distribution can be accomplished easily through importance sampling. (For an overview see Train (2009) pages 210-211.)

Integrating over $\boldsymbol{\omega}_i$ is straightforward as we can simply draw ω_i from its normal distribution given Θ. We draw censored values $\tilde{\theta}_{it}^k < 0$ period by period. If peak usage is censored and off-peak usage is positive, we draw $\tilde{\theta}_{it,b}^{pk}$ from $f^{\theta,2}(\tilde{\theta}_{it,b}^{pk} \mid \tilde{\theta}_{it,b}^{pk} < 0, \tilde{\theta}_{it,b}^{op}, \boldsymbol{\mu}_{i,b})$, where the density $f^{\theta,2}$ represents the truncated univariate normal density of θ_{it}^{pk} conditional on θ_{it}^{op}, 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_{it}^{pk} and θ_{it}^{op} from a truncated bivariate normal distribution. As for s_{it} we draw from truncated normal distributions via importance sampling.

Note that we write L_{it} conditional on a simulated draw of the firm consideration set, F_{itb}. We do so because we substitute the conditional plan choice probability $\Pr\left(j_{it} \mid \Im_{it}, F_{itb}\right)$ evaluated at a randomly drawn $F_{itb} \in \left\{F_{it}^1, F_{it}^2, F_{it}^3\right\}$ in place of the unconditional plan choice probability $\Pr\left(j_{it} \mid \Im_{it}\right)$ from equation (30) when calculating L_{it} from equation (32) at a particular simulation draw b. This approach reduces computation time because it means that we only have to compute the expected utility for a third of outside firm plans at each simulation draw.

D.4.3 Computing Bounds

We now turn to the computation of the bounds $[\underline{s}_{it}, \bar{s}_{it}]$ from which we draw simulated values of s_{it}. First consider the case in which all plans in i's choice set include free off-peak calling. In this case only the consumer's beliefs about θ_{it}^{pk} and $r_{it}^{in,pk}$ (rather than also those about θ_{it}^{op} and $r_{it}^{in,op}$) affect plan expected utility. Let $\tilde{\mu}_{\theta it}^{pk} = \tilde{E}[\tilde{\theta}_{it}^{pk} \mid s_{it}]$ and $(\tilde{\sigma}_{\theta t}^{pk})^2 = \tilde{V}[\tilde{\theta}_{it}^{pk} \mid s_{it}]$, for which formulas are given in Appendix B.3. Our approach is to first calculate bounds on $\tilde{\mu}_{\theta it}^{pk}$ that are implied by plan choice j_{it}, and then invert $\tilde{\mu}_{\theta it}^{pk} = \tilde{\mu}_{it}^{pk} + \rho_{s,pk}\sigma_{\varepsilon}^{pk} s_{it}$, to find corresponding bounds on s_{it}. In particular, $\tilde{\mu}_{\theta it}^{pk} \in [\underline{\mu}_{\theta it}, \bar{\mu}_{\theta it}]$ implies that

$$\frac{\underline{\mu}_{\theta it} - \tilde{\mu}_{it}^{pk}}{\sigma_{\varepsilon}^{pk}\rho_{s,pk}} \leq s_{it} \leq \frac{\overline{\mu}_{\theta it} - \tilde{\mu}_{it}^{pk}}{\sigma_{\varepsilon}^{pk}\rho_{s,pk}}. \tag{37}$$

Expected utility for a plan with free off-peak calling varies across individuals and time depending on the three parameters $\{\tilde{\mu}_{\theta,it}^{pk}, \tilde{\sigma}_{\theta t}^{pk}, \alpha_i^{pk}\}$. To calculate the bounds $\tilde{\mu}_{\theta it}^{pk} \in [\underline{\mu}_{\theta it}, \bar{\mu}_{\theta it}]$, we begin by calculating expected utility of each university plan on a three dimensional grid of $\tilde{\mu}_{\theta,it}^{pk}$, $\tilde{\sigma}_{\theta t}^{pk}$, and α_i^{pk} values. (This computation occurs only once, prior to evaluating the likelihood.) The $\tilde{\mu}_{\theta,it}$ grid is regular between -100 and 1200 minutes, and each point is spaced 10 minutes apart. The $\tilde{\sigma}_{\theta,t}^2$ grid is irregular, where the grid points are $\{\tilde{\sigma}_{\theta,1}^2, ..., \tilde{\sigma}_{\theta,T}^2\}$. The α_i^{pk} grid has 11 equally spaced points between 0 and 1.

Next, for each consumer i and period t, we fix $\tilde{\sigma}_{\theta t}^{pk}$ and a simulated α_{ib}^{pk} and (using linear interpolation between adjacent α_i^{pk} grid points) compute expected university plan utilities over a one dimensional grid of $\tilde{\mu}_{\theta,it}^{pk}$ values. At each $\tilde{\mu}_{\theta,it}^{pk}$ value on the grid, we determine the optimal university plan. Then bounds on $\tilde{\mu}_{\theta,it}$ implied by observed plan choice j_{it} are computed using linear interpolation. For example, if plan 0 is chosen, and plan 0 is optimal for all $\tilde{\mu}_{\theta,it}$ grid points $\{1, ..., k\}$, and plan 1 is optimal for point $k+1$, then we assume that the upper bound on $\tilde{\mu}_{\theta,it}$ occurs where the interpolated utilities cross for plans 0 and 1 in between points k and $k+1$.[5] Bounds on signals immediately follow: $[\underline{s}_{it}, \bar{s}_{it}] = \left([\underline{\mu}_{\theta it}, \bar{\mu}_{\theta it}] - \tilde{\mu}_{it}^{pk}\right) / \left(\sigma_{\varepsilon}^{pk}\rho_{s,pk}\right)$.

The preceding algorithm works when off-peak calling is free. At some dates, however, plan 0 charges 11 cents per minute off-peak as well as on peak. We refer to this pricing as the *costly*

[5] We did experiment with finding where actual utilities crossed, rather than interpolated utilities, in a simplified version of the model where we estimated beliefs from period 1 choices in the fall of 2002. We found that estimated beliefs were the same for each method, but the interpolation was much faster.

version of plan 0. The expected utility of costly plan 0 varies across individuals and time due to variation in $\{\tilde{\mu}^{op}_{\theta,it}, \alpha^{op}_i\}$ as well as $\{\tilde{\mu}^{pk}_{\theta,it}, \tilde{\sigma}^{pk}_{\theta t}, \alpha^{pk}_i\}$. (The parameter $\tilde{\sigma}^{op}_{\theta}$ is constant across time because there is no learning about off-peak type.) To account for this variation, we calculate the difference in expected utility between free-off peak calling and 11 cent per minute off-peak calling on a two dimensional grid of $\{\tilde{\mu}^{op}_{\theta,it}, \alpha^{op}_i\}$ values.

Next, we notice that for a particular i and t, $\tilde{\mu}^{op}_{\theta,it}$ is a function of $\tilde{\mu}^{pk}_{\theta,it}$:

$$\tilde{\mu}^{op}_{\theta,it} = \mu^{op}_i + \rho_{s,op}\sigma^{op}_{\varepsilon} s_{it} = \mu^{op}_i + \sigma^{op}_{\varepsilon}\frac{\rho_{s,op}}{\rho_{s,pk}}\left(\tilde{\mu}^{pk}_{\theta,it} - \tilde{\mu}^{pk}_{it}\right).$$

Therefore, for each i and t we can compute (via linear interpolation) the expected utility of the costly version of plan 0 over a grid of $\tilde{\mu}^{pk}_{\theta,it}$ values, conditional on $\tilde{\sigma}^{pk}_{\theta t}$ and simulated values α^{pk}_{ib} and α^{op}_{ib}. We then compute bounds as before.

Finally, note that the bounds $\underline{s}_{it}$ and $\bar{s}_{it}$ are functions of the time t plan choice j_{it} and information set $\Im_{itb}$ (excluding s_{itb}). In particular, period t bounds depend on the previous period's signal $s_{i,t-1,b}$ and latent taste shock $\tilde{\theta}^{pk}_{i,t-1,b}$ (via their effect on $\tilde{\mu}^{pk}_{\theta,it}$). Therefore we must first draw signals $\mathbf{s}^{t-1}_{ib}$ and censored shocks $\tilde{\boldsymbol{\theta}}^{-,t-1}_{ib}$ in order to compute bounds $\underline{s}_{it}$ and $\bar{s}_{it}$ and draw a signal s_{itb}.

D.5 Computational Procedures

When we compute our standard errors, each time we run our estimator on a different subsampled data set we also use a different set of pseudo random draws to compute the log likelihood function. Our standard errors will therefore also account for simulation error.

We wrote the program to evaluate the likelihood in R and Fortran. The evaluation of the 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 $\alpha^{9,op}_i$ 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 $\alpha^{9,op}_i$'s, and the outer loop maximizes the likelihood.

We summarize the algorithm for computing these variables in four steps. Step 1 is to compute $\alpha^{op,9}_{i,b}$ 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 b, we compute $\alpha^{op,9}_{i,b}$ as the solution to equation (22) in Appendix C, which extends equation (9) to account for in-network calling. As this equation does not have an analytic solution, we compute $\alpha^{op,9}_{i,b}$ with a nonlinear equation solver. The result of this step is used to compute the structural error for $r^{op,9}_{it}$.

The next three steps compute the calling threshold vector $\mathbf{v}^*_{it,b}$ and $\tilde{\boldsymbol{\theta}}_{it,b}$ period-by-period. Because the $\mathbf{v}^*_{it,b}$ is a function of past values of $\tilde{\boldsymbol{\theta}}_{it,b}$ through the Bayesian learning, 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 III.D. First, consumer beliefs about μ^{pk}_i, $(\tilde{\mu}^{pk}_{it,b}, \tilde{\boldsymbol{\sigma}}^2_{it})$ are updated via Bayes rule. Second, beliefs about $\tilde{\boldsymbol{\theta}}_{it,b}$ are computed from $(\tilde{\mu}^{pk}_{it,b}, \tilde{\boldsymbol{\sigma}}^2_{it})$ and $\mu^{op}_{i,b}$. (No updating is required for $t = 1$.) In step 3 we calculate $\mathbf{v}^*_{it,b}$ following it's characterization in Appendix C, which depends on the beliefs calculated in step 2. Recall that 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 (20) or (21) which are the extensions to equation (4) that account for in-network calling given in Appendix C). In step 4, we calculate $\tilde{\boldsymbol{\theta}}_{it,b}$. When $\boldsymbol{\theta}_{it,b}$ is not censored, we can compute $\tilde{\boldsymbol{\theta}}_{it,b}$ from observed usage conditional on β and $\mathbf{v}^*_{it,b}$ using equations (34)-(35) in Appendix C. When censoring occurs, we use the simulated value for $\tilde{\boldsymbol{\theta}}_{it,b}$.

With $\alpha^{op,9}_{i,b}$, $\tilde{\boldsymbol{\theta}}_{it,b}$ and $\mathbf{v}^*_{it,b}$ in hand we can compute the choice probabilities and the density of observed usage in equation (23). Choice probabilities are calculated from consumer expected utilities, $U_{itj,b}$, which are defined in equation (19) for all plans in consumers' choice sets. These depend on plan-specific calling threshold vectors $\mathbf{v}^*_{itj,b}$ (which are all computed as part of the vector $\mathbf{v}^*_{it,b}$). Next, notice that

$$V\left(q(v^k_{itj}, x^k_{it}), x^k_{it}\right) = x^k_{it}\frac{1}{\beta}\hat{q}(v^k_{itj})\left(1 - \frac{1}{2}\hat{q}(v^k_{itj})\right) \tag{38}$$

is linear in x^k_{it} for $k \in \{\text{pk-out, pk-in, op-out, op-in}\}$ and hence $\sum_k V(q^k_{it}, x^k_{it})$ is linear in θ^{pk}_{it} and θ^{op}_{it}. Thus $E\left[\sum_k V(q^k_{it}, x^k_{it})\right]$ can be computed analytically (up to evaluation of the standard normal cumulative distribution). Moreover, the expected price $E\left[P(q)\right]$ is a linear function of the expected amount θ^{pk}_{it} exceeds $Q_{ijt}/\hat{q}(v^{pk}_{itj,b})$ (or $x^{pk,out}_{it}$ exceeds $Q_{ijt}/\hat{q}(v^{pk,out}_{itj,b})$ for free-in-network), which we can also evaluate analytically in all cases except for when plan 2 offers free in-network minutes. In the latter case we approximate the expectation with Gaussian quadrature.

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.[6] Conditional on these choices of usage parameters, we choose initial belief parameters to

[6]We assume that v^{pk}_{itj} is equal to 3 cents for plan 3, 5 cents for plan 2, and 8 cents for plan 1, and maximize

match the observed plan shares. To do so, 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.

E Equilibrium Price Calculation

To compute the pricing equilibrium we begin with 1,000 simulated consumers who we assume are in the market for $T = 12$ periods, and 3 identical firms offering 3 plans (we assume that 4 plans are offered in the calibration). In period 1, simulated consumers choose from the set of all plans offered by any of the three firms. In later periods, the consumer chooses between the plans offered by the firm they chose in period $t-1$, those offered by a randomly assigned alternative firm, and the outside option, consistent with the model we estimate. We denote the firm's fixed cost of serving a customer as FC and the marginal cost of providing peak minutes as c (off-peak minutes are assumed to have zero marginal cost).

We solve for a symmetric equilibrium by iterating on firm best response functions: we start by assuming all firms offer a set of plans that looks close to what is offered in the data, and compute one firm's best response to that using a numerical optimizer. We then assume that all firms offer the best response plans, and compute the best response to those plans. The algorithm converges when the difference in prices between one best response and the next are below a threshold (we take the average percentage change in the characteristics and use a tolerance of 1 percent).

The remainder of this section provides details on how the best responses are calculated. For $t > 1$, a simulated consumer i is assigned a vector of draws on the type, signals, tastes, firm errors, and plan consideration $(\boldsymbol{\omega}_i, \boldsymbol{s}_i, \boldsymbol{\theta}_i, \boldsymbol{\eta}_i, \boldsymbol{PC}_i)$. In the first period, however, for each possible plan choice we draw a signal conditional on being inside bounds corresponding to that plan choice and then integrate over the firm logit error η_f to compute a choice probability for the plan. This different treatment in the first period reduces the discontinuities in the simulated profit function that arise from simulating firm errors $\boldsymbol{\eta}_{i1}$ and a single signal s_{i1}. Moreover, doing this only for $t = 1$ avoids the computational problem that the number of possible plan choice vectors $\{j_{i1}, ..., j_{iT}\}$ increases exponentially with T.

the likelihood of usage conditional on those guesses at v^{pk}_{itj}. We chose those values of v^{pk}_{itj} because they matched the average values of v^{pk}_{itj} 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^{pk}_{itj} to arrive at starting points; we solve for the endogenous v^{pk}_{itj}'s when running the full simulated maximum likelihood.

Our approach is to begin by solving for prices for the $T = 1$ equilibrium. This is a smooth problem for which we use a derivative-based BFGS optimizer. We then solve the $T = 12$ problem using the solution for the $T = 1$ problem as a starting point. This is not a smooth problem and hence we use a simplex optimizer to solve for the best response.

During optimization we put constraints on the plan characteristics the firm can offer. For all plans j offered by the same firm, the constraints are:

$$\begin{aligned} M_j &\geq 0; \quad M_j \geq M_{j-1} \\ Q_j &\geq 0; \quad Q_j \geq Q_{j-1} \\ 0 &\leq p_j \leq \overline{p}; \quad p_j \leq p_{j-1}. \end{aligned}$$

These constraints ensure we produce plans that look like what we observe in the data. Typically we find an interior solution so the constraints do not bind - exceptions to this are the upper bound $\bar{p}$ on overage rates without bill-shock regulation and the lower bound $Q_j \geq 0$ when we de-bias consumers. To ensure the optimizer does not wander into bad areas of the parameter space, we take a penalty function approach.

F Calibration Exercise

As discussed in Section VII, to run our endogenous-price counter-factual simulations we first calibrate the logit-error weight $1/\lambda$ and the marginal cost of providing a peak call, c. For the exercise and all the counterfactuals we assume that the fixed cost of serving a customer, FC, is \$15.00. Our algorithm for the calibration 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. (Sprint did not offer any local plans, so we did not include them.) 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, we solve for equilibrium prices and compare them to prices observed over the two year period. When computing equlibrium prices, we restrict included minute allowances, Q_j, to weakly increase with fixed fees, M_j, and bound overage rates $p_j \leq 50¢$. Formally, firm j offers $k = 1, .., 4$ plans, where the plans are ordered according to their fixed fees, and we denote the predicted price schedule as a function of λ and c as $(\hat{M}_{jk}(\lambda, c), \hat{p}_{jk}(\lambda, c), \hat{Q}_{jk}(\lambda, c))$ and the observed price schedule of a firm during month t as $(M_{jkt}, p_{jkt}, Q_{jkt})$. We choose λ and c to minimize the squared difference between observed and predicted prices:

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

This calibration exercise yields $\hat{\lambda} = 0.03$ and $\hat{c} = 0.02$. We used a grid search to find the optimal $\hat{\lambda}$ and $\hat{c}$.

After the calibration we tested how sensitive our parameter estimates from our primary specification ($\lambda = 1$) differed for two alternate specifications, 1) a specification which fixed $O = 0$ and let λ be free, and 2) a specification where λ was fixed to its calibrated value of 0.03 and the rest of the parameters were free. Specification 1) yielded a value of λ of roughly 0.06. Apart from λ, the difference between the estimated parameters in the primary specification and the alternatives were negligible. (To make sure the small differences were not due to an optimization issue we redid the estimations with several different step sizes, ranging from large to very small. The optimizer converged to the same estimates for all step sizes we chose.)

G Robustness

G.1 Overage Rate

A natural concern is that our results will be sensitive to the bound we impose on the overage rate since that effects how far overage rates fall when bill-shock regulation is imposed. To address this issue, Table 15 reports simulations with \$0.35 and \$0.75 overage rate caps. (For context, empirically observed overage rates range from \$0.35 to \$0.50 in our sample period.) We find that our results are robust to the alternative overage rate limits.

The intuition for robustness is as follows: First, profits are unaffected by the maximum overage rate as fixed fees adjust in equilibrium to compensate. Second, average usage varies by less than 10 minutes per month between \$0.35 and \$0.75 caps because (1) overconfidence diminishes the effect of the overage rate on calling thresholds, and (2) minute allowances adjust in equilibrium in an off-setting manner. Third, social surplus changes little because calling changes little. Fourth, consumer surplus changes little because it is the difference between two stable quantities, social surplus and profits.

G.2 Learning

Columns 4 and 5 of Table 5 show that our predictions about the consequences of bill-shock regulation depend importantly on estimated biases. A natural question, then, is how much the predictions

Table 15: Effect of Bill-Shock Regulation for Different Overage Rate Caps

Overage rate cap:		50 cent cap		35 cent cap		75 cent cap	
Regulation:		None	Bill shock	None	Bill shock	None	Bill shock
		(1)	(2)	(3)	(4)	(5)	(6)
Plan 1	M	42.88	39.28	46.85	39.28	37.18	39.28
	Q	216	0	188	0	225	0
	p	0.50	0.17	0.35	0.17	0.75	0.17
	Share	39	26	36	26	35	26
Plan 2	M	48.64	50.66	52.73	50.66	41.33	50.66
	Q	383	80	365	80	363	80
	p	0.50	0.12	0.35	0.12	0.75	0.12
	Share	38	23	36	23	40	23
Plan 3	M	58.12	68.23	60.82	68.23	50.98	68.23
	Q	623	540	613	540	576	540
	p	0.50	0.12	0.35	0.12	0.75	0.12
	Share	14	40	17	39	17	39
Outside good share		10	12	11	12	9	12
Usage		240	239	245	240	236	240
Overage revenue		223	152	178	153	294	153
Annual profit		501	509	501	508	508	508
Annual consumer welfare		903	870	911	884	916	884
Annual total welfare		1404	1379	1413	1393	1424	1393
Δ Annual profit			7		7		1
Δ Annual consumer welfare			-33		-27		-32
Δ Annual total welfare			-26		-20		-31

All welfare and profit numbers are expressed in dollars per customer per year. We simulate 10,000 consumers for 12 months.

would change if time passes allowing consumer beliefs to evolve via learning but all else remains equal. To answer this question, we ran our bill-shock counterfactual simulation after allowing consumers beliefs to evolve for either 12 months or 36 months before making their first plan choice. Results are shown in Table 16. One would expect the magnitude of welfare changes in response to bill-shock regulation to decline with learning as we have already shown that they disappear completely with de-biasing. This is what we find. Predicted annual consumer surplus losses from bill-shock regulation decline from \$33 in our primary counterfactual to \$27 with 12 months pre-learning and to \$12 with 36 months pre-learning. The qualitative prediction that average consumer surplus falls a small amount is robust. The most notable difference in pricing is that the plan menu collapses to two plans with bill-shock regulation given additional learning. This response to additional learning makes sense because our main specification already shows that de-biasing reduces the menu to two tariffs.

G.3 Price Sensitivity

As noted in footnote 27, one might be concerned that our estimate of β is biased. On the one hand, our estimate of β might be too high if (in contrast to our assumption) some of the increase in usage at 9pm occurs because consumers delay calls until after 9pm. On the other hand, our estimate of β is much smaller than what would be estimated from the jump in usage at 9pm alone. We therefore perform sensitivity analysis constraining β to alternative values.

To save computational time, we use a half-sample of the data. Table 17 reports estimated parameters from three specifications. Specification (1) estimates our full model on the half-sample. The estimate of β is 2.86, slightly higher than the estimate on the full sample of 2.70. Specification (2) holds fixed β at half its estimated value and re-estimates the other 27 parameters. Likewise, specification (3) holds fixed β at 1.5 times its estimated value and re-estimates the other 27 parameters. (Note that these alternative values of β lie more than 2 standard errors away from our estimate. Thus they are meaningfully different values.)

As expected, many parameters are stable across the three specifications. For instance, off-peak parameters, in-network calling parameters, near-9pm calling parameters, the plan consideration probability P_C, overconfidence δ, and correlations $\rho_{s,pk}$, ρ_{μ}, $\rho_{\tilde{\mu},op}$, and ρ_{ε}, are all stable. Additionally, if one uses the standard errors estimated in the main specification as a baseline, all but μ^{op} for the large value of β are within 2 standard deviations of the baseline estimates. Also as expected, other parameters vary monotonically with β in reassuring ways. As β increases, larger values of θ are required to explain the same usage when marginal price is positive. Thus it is reassuring that $\tilde{\mu}_0^{pk}$, $\tilde{\sigma}_{\mu}^{pk}$, $\sigma_{\varepsilon}^{pk}$, and μ_0^{pk} increase with β. (In contrast it is not surprising that μ_0^{op} and other off-peak

Table 16: Effect of Bill-Shock Regulation After 12 or 36 Months of Consumer Learning

Learning:		None		12 months		36 months	
Regulation:		None	Bill shock	None	Bill shock	None	Bill shock
		(1)	(2)	(3)	(4)	(5)	(6)
Plan 1	M	42.88	39.28	42.6	49.16	43.98	58.05
	Q	216	0	204	97	207	280
	p	0.50	0.17	0.50	0.15	0.50	0.16
	Share	39	26	22	40	19	56
Plan 2	M	48.64	50.66	45.81	63.91	47.12	63.14
	Q	383	80	331	441	323	478
	p	0.50	0.12	0.50	0.15	0.50	0.16
	Share	38	23	43	48	38	31
Plan 3	M	58.12	68.23	51.94		52	
	Q	623	540	497		479	
	p	0.50	0.12	0.50		0.50	
	Share	14	40	25		33	
Outside good share		10	12	10	12	10	13
Usage		240	239	247	245	253	261
Overage revenue		223	152	215	102	199	65
Annual profit		501	509	485	481	483	471
Annual consumer welfare		903	870	948	921	955	943
Annual total welfare		1404	1379	1432	1402	1438	1414
Δ Annual profit			7		-3		-12
Δ Annual consumer welfare			-33		-27		-12
Δ Annual total welfare			-26		-30		-24

All welfare and profit numbers are expressed in dollars per customer per year. We simulate 10,000 consumers learning for 12 or 36 months to generate initial beliefs, followed by 12 months of plan and usage choices for pricing and welfare calculations.

Table 17: Parameter Estimates at Different β Values

Parameter	Est. β (1)	Low β (-50%) (2)	High β ($+50\%$) (3)
β	2.86	1.43	4.29
$\tilde{\mu}_0^{pk}$	219.97	205.28	274.99
μ_0^{pk}	273.69	257.51	311.76
μ_0^{op}	402.48	406.94	417.24
$\tilde{\sigma}_\mu^{pk}$	195.51	168.67	202.73
σ_μ^{pk}	146.98	160.32	160.07
σ_μ^{op}	431.71	453.14	425.8
$\rho_{\tilde{\mu},pk}$	0.22	0.43	0.17
$\rho_{\tilde{\mu},op}$	0.01	0.01	0.01
ρ_μ	0.59	0.62	0.66
σ_ε^{pk}	254.76	230.55	296.01
σ_ε^{op}	344.08	340.91	351.04
ρ_ε	0.54	0.53	0.54
$\rho_{s,pk}$	0.45	0.43	0.44
δ	0.36	0.36	0.37
$\mu_\alpha^{pk,9}$	-0.01	-0.01	-0.01
$\sigma_\alpha^{pk,9}$	0.06	0.06	0.06
$\sigma_e^{pk,9}$	0.11	0.12	0.11
$\sigma_e^{op,9}$	0.12	0.12	0.12
P_C	0.06	0.06	0.06
O	1.16	1.83	0.84
μ_α^{pk}	0.33	0.36	0.31
μ_α^{op}	0.39	0.4	0.39
σ_α^{pk}	0.18	0.18	0.18
σ_α^{op}	0.17	0.17	0.17
σ_e^{pk}	0.18	0.19	0.18
σ_e^{op}	0.17	0.17	0.16
δ_r	0	0	0
b_1	-53.72	-52.23	-36.78
b_2	0.84	0.59	0.87
Log-likelihood	142680	142733	142763

Estimation uses a half-sample of the data.

parameters do not increase with β as off-peak usage is typically free and independent of β.)

Several parameters are worth commenting on in more detail. Aggregate mean bias b_1 appears stable to reducing β but declines in absolute value for larger β (although it is still negative). Interestingly, the correlation $\rho_{\tilde{\mu},pk}$ declines with β which means (following equation (8) that conditional mean bias b_2 increases with β. Thus if we have overestimated β we may also have slightly overestimated conditional mean bias. The intuition for this effect is that large plan choice is correlated with high usage. As β increases, the model attributes more of this difference to the price effect and less to the selection of high users into large plans. The model turns down self-selection by reducing the correlation between true types and point estimates so that consumers' plan choices are increasingly noisy.

Next we conduct our counterfactual simulation of the effect of bill-shock regulation at estimates corresponding to the three different values of β. These are reported in Table 18. First note that the outside good is smaller and more stable under regulation at lower values of β, which makes sense since the value of phone service relative to the outside good is higher at lower β. Second note that overage rates fall more with regulation when β is higher, which again makes sense because larger marginal price cuts are needed for more price sensitive consumers. Third, included minute allowances increase with β for similar reasons. Fourth, usage increases with β. This might first sound surprising, that more price sensitive consumers consume more. However, note that the change in usage depends both on varying β and the accompanying changes in the parameters governing the distribution of θ. Since the estimates are all trying to fit the same usage levels for the same prices in the demand data we should expect that at equal prices usage would not vary with the specifications. So it should actually not be a surprise that usage increases when marginal price decreases at larger β values.

Turning to the welfare predictions, our result that consumer surplus and social welfare on average decrease by a small amount is robust for different values of β. For some values of β we find profits increase and at others we find profits decrease. This finding is partially (but not completely) explained by the fact that at higher β the outside good is a closer substitute and firms have less market power to raise fixed fees following regulation. More generally, it may reflect that fact that the regulation does not have a strong direct effect on market power and hence there is not a clear prediction a priori about whether profits should go up or down.

G.4 Heterogeneity

As our sample consists entirely of students a possible concern is that our sample is more homogeneous and more or less price sensitive than the general population. To diagnose whether this

Table 18: Effect of Bill-Shock Regulation at Alternate β Values.

Price sensitivity:		$\beta = 2.86$ (Est.)		$\beta = 1.43$ (-50%)		$\beta = 4.29$($+50\%$)	
Regulation:		None	Bill shock	None	Bill shock	None	Bill shock
		(1)	(2)	(3)	(4)	(5)	(6)
Plan 1	M	44.87	41.1	40.88	27.32	45.05	47.3
	Q	216	0	169	0	281	0
	p	0.5	0.15	0.5	0.3	0.5	0.11
	Share	35	31	34	29	35	28
Plan 2	M	49.28	61.72	48.22	58.47	48.8	65.52
	Q	364	307	325	264	438	377
	p	0.5	0.15	0.5	0.29	0.5	0.11
	Share	34	37	41	38	34	35
Plan 3	M	56.02	69.31	59.15	69.02	55.24	71.09
	Q	564	627	540	543	663	731
	p	0.5	0.15	0.5	0.3	0.5	0.11
	Share	19	18	20	28	14	16
Outside good share		12	14	4	5	17	22
Usage		244	236	237	227	274	270
Overage revenue		222	139	251	214	216	107
Annual profit		505	494	562	575	458	457
Annual consumer welfare		779	760	2298	2255	358	321
Annual total welfare		1284	1255	2861	2831	816	778
Δ Annual profit			-11		13		-1
Δ Annual consumer welfare			-18		-43		-37
Δ Annual total welfare			-29		-30		-38

All welfare and profit numbers are expressed in dollars per customer per year. We use parameter estimates from a half-sample of the data. We simulate 10,000 consumers for 12 months.

may be true, we use our data to estimate a version of Jiang's (2013) model adapted to our data. Comparison of the resulting estimates to those reported in Jiang (2013) that are based on national data suggest that students in our sample may indeed be more homogeneous and less price sensitive than the general population. (Details of this exercise following in Appendix H.)

We therefore conduct a sensitivity analysis to the variance of the joint normal distribution of consumer type $\{\tilde{\mu}_{i1}^{pk}, \mu_i^{pk}, \mu_i^{op}\}$ in conjunction with higher values of β. We perform our bill shock exercise with two modified sets of estimates. In one set of estimates we scale up σ_μ^{pk} by 2 and set $\beta = 4$, and in the second we scale σ_μ^{pk} up by 2.5 and set β to 1.5 times its estimated value. To preserve the biases, we rescale $\tilde{\sigma}_\mu^{pk}$ by the same factor as σ_μ^{pk}, and keep all the model's correlation parameters the same (so the covariances are also multiplied by 2 or 2.5, depending on the alternative specification). The results of this exercise are shown in columns 3 through 6 of Table 19 (columns 1 and 2 replicate columns 1 and 3 of Table 5 in the main body of the paper, showing the predicted impact of bill shock regulation using our estimated parameters).

Without bill shock, both fixed fees and the number of included minutes in a plan increase as the variance of μ_i^{pk} and the price coefficient rises. As the variance in the population rises there are more consumers who place a high value on talking, which likely explains both the rise in fixed fees and the number of included minutes. Similarly, there are more consumers who place a low value on talking, which likely explains the rise in the share of the outside good. The larger shares of the outside good also reflect higher values of β.

The impact of bill-shock regulation on prices is similar to what we find in the main counterfactual: fixed fees rise (except for plan 1), the number of included minutes drops, and the overage rates drop. Turning to the welfare effects of the counterfactuals, they are similar in that consumer welfare drops, and total welfare drops. One difference relative to our main results is that firm profits also drop, which is likely a result of higher consumer price sensitivity. In conclusion, we believe this exercise suggests that the qualitative implications of our paper would be unchanged if our sample was more heterogeneous and more price sensitive.

H Replication of Jiang's (2013) Model Using Our Data

In this section we estimate Jiang's (2013) model on our data. We attempt to follow Jiang's (2013) approach as closely as possible, given the differences between our two datasets. The version of Jiang's (2013) model that we estimate has 7 parameters: we estimate the population average of the taste parameter (θ_i), the standard deviation of the taste parameter, the standard deviation of the perception error (ω), an overage price coefficient, a fixed-fee price coefficient, and two dummy

Table 19: Effect of Bill-Shock Regulation for Populations that are More Heterogeneous and More Price Sensitive

Scenario:		Our estimates		$2 \times \sigma_\mu^{pk}$, $\beta = 4$		$2.5 \times \sigma_\mu^{pk}$, $\beta = 1.5 \times \hat{\beta}$	
Regulation:		None	Bill shock	None	Bill shock	None	Bill shock
		(1)	(2)	(3)	(4)	(5)	(6)
Plan 1	M	42.88	39.28	46.74	44.91	48	43.96
	Q	216	0	232	7	242	9
	p	0.5	0.17	0.5	0.12	0.5	0.13
	Share	39	25	39	37	37	36
Plan 2	M	48.64	50.66	54.46	66.32	55.74	65.92
	Q	383	80	531	504	582	520
	p	0.5	0.12	0.5	0.13	0.5	0.13
	Share	38	23	31	27	28	23
Plan 3	M	58.12	68.23	66.84	72.1	68.49	72.56
	Q	623	540	∞	∞	∞	∞
	p	0.5	0.12	N/A	N/A	N/A	N/A
	Share	13	40	13	17	15	20
Outside good share		10	12	18	20	20	21
Usage		240	240	263	267	280	273
Overage revenue		223	152	161	112	145	102
Annual profit		501	509	453	452	441	430
Annual consumer welfare		903	870	408	400	305	304
Annual total welfare		1404	1379	861	852	745	734
Δ Annual profit			7		-1		-11
Δ Annual consumer welfare			-33		-8		-1
Δ Annual total welfare			-26		-9		-12

All welfare and profit numbers are expressed in dollars per customer per year. We simulate 10,000 consumers for 12 months.

variables for plan characteristics - whether a plan includes free off-peak and whether a plan included free in-network. The last two parameters were included to ensure our approach is as comparable as possible to Jiang's (2013) model, which also includes dummy variables for plan characteristics (although the included characteristics such as free long-distance and free roaming indicators are different than ours). Jiang (2013) also includes demographic interactions in the price coefficients that we cannot include because we do not have demographic information. However, her demographic variables are quite coarse and it may be sufficient to assume they are zero for our sample - they include a dummy variable for whether an individual is in a family, a dummy for whether the individual is high-income, and a dummy for whether the individual is over 55 or not. Jiang's (2013) estimates of the demographic interactions are not statistically significant. Jiang (2013) also estimates the variance of the logit error in her specification. We found we could not identify the logit error variance in our specification so we fixed the standard deviation of the logit error to the value she estimates - 66.94. We assume that consumers are rational, as Jiang (2013) does, and restrict the mean of the perception error to be 1.

Following Jiang (2013) we estimate the model parameters using method of moments.[7] Our moment selection follows hers very closely. We break our sample up into 9 periods (our period definitions correspond to the periods outlined in Appendix A Table 6; we drop the initial period 8/02-10/02 where plan 0 was unavailable), and choose parameters to match the following moments:

1. Share of plan j in period t.
2. Average probability of being 10 percent below or 10 percent above the number of included minutes in period t (for 3 part tariffs only).
3. Average usage in period t.
4. Coefficient of variation in usage in period t.
5. Correlation between usage and overage rate in period t.

We construct our moments using only the first observed bill for each consumer because Jiang's (2013) model does not incorporate plan switching or grandfathering of plans, and because we believe that Jiang's (2013) sample selection procedure will oversample new customers. In terms of weighting, moments are weighted according to 1) the standard deviation of the outcome variable,

[7]It would not in fact be possible to estimate Jiang's (2013) model using maximum likelihood because Jiang's (2013) model assumes that all consumers use a strictly positive number of minutes every month. Her model would predict zero likelihood for observations in our data where we observe zero usage.

Table 20: Estimates of Jiang's (2013) model on our data

Parameter	Estimate from our data		Estimate from Jiang (2013)		P-Value of Col 1 - Col 2
Mean Preference Parameter	5.12	(0.08)	4.93	(0.6)	0.76
Standard Deviation of Unobserved Heterogeneity	0.42	(0.14)	0.87	(0.21)	0.07
Standard Deviation of Log of Perception Error	0.81	(0.05)	0.33	(0.29)	0.11
Price Coefficient with Respect to Overage Price	-8.98	(2.15)	-19.8	(56.96)	0.85
Price Coefficient with Respect to Fixed Fee	-6.9	(1.66)	-22.6	(0.54)	0
Expensive Off-Peak	-160.64	(55.86)	-	-	-
Free In-Network	52.77	(58.21)	-	-	-

The first column of this table shows the coefficient estimates obtained when our implementation of Jiang's (2013) model is estimated on our dataset. Standard errors are shown in parentheses. The second column show estimates of Jiang's (2013) model on Jiang's (2013) data, from Table 4, page 25 of her paper. The corresponding standard errors are shown in parentheses. The final column shows the p-value of the hypothesis test that our estimate is equal to Jiang's (2013) estimate, assuming our data sets are independent.

and 2) the proportion of sign-ups occurring in each period. Item 1) ensures that moments with large outcome variables (such as minutes usage) are not overweighted in the objective function relative to moments such as market shares, and item 2) puts more weight on moments which are measured more precisely. To construct our predicted moments, we simulate plan choices and usage for 100 consumers in each period.

Our estimates are shown in Table 20. The first column of the table shows the parameter estimates arising from our implementation of Jiang's (2013) model on our data, with standard errors in parentheses. The estimates of comparable parameters of Jiang's (2013) model applied to Jiang's (2013) data are shown in the second column (they are copied from Jiang (2013), Table 4, page 25). The third column shows the p-value of the hypothesis test that the parameter estimate in column 1 equals that of column 2, assuming that the two parameter estimates are independently distributed (equivalently, assuming that the variables in our dataset are statistically independent from those in Jiang's (2013) dataset).

Turning to the estimated parameters, the mean taste parameters are similar across both specifications. Our estimate of the standard deviation of θ_i is 0.42, while Jiang's (2013) is much larger at 0.87. If we interpret differences in the parameter estimates due to differences in the samples used for estimation (as opposed to differences in specification, or model implementation), then the estimates suggest that our sample is more homogenous than Jiang's (2013) sample, which is unsurprising since our sample consists of students while hers is drawn from the general population. This conclusion must be tempered somewhat due to the large standard error in Jiang's (2013) estimate of the standard deviation of θ_i - as shown in the third column of the table the two parameters are not statistically different at the 5 percent level. Our estimate of the perception error variance is 0.81, which is more than twice Jiang's (2013) estimate of 0.33. Again, the standard error around her estimate of this standard deviation is 0.29, indicating that our estimate is not statistically different than hers. Turning to the overage price coefficient, our estimate is much smaller than Jiang's (2013) estimate. However, the standard error around Jiang's (2013) estimate of this coefficient is extremely large at 57, so our estimate of the price coefficient is not statistically different from hers. The only parameter that is statistically different from Jiang's (2013) is the price coefficient on the overage price, which is around one third the magnitude. This suggests that our sample is less price sensitive than Jiang's (2013) sample which is reasonable - students likely come from higher income families relative to the general population so it makes sense they would be less sensitive to prices.

We next turn to the question of how our bill-shock counterfactual might be affected if the estimates from the model presented in the body of the paper are biased in a manner that is similar to the differences we observe between the two parameter estimates in Table 20. Because the two models are very different this exercise is imperfect. For example, there is no counterpart to the perception error in the model in our paper, so it is not clear to us what the implications of having a larger perception error variance are. Similarly, as Jiang's (2013) dataset is a cross-section there is no counterpart to our ϵ_{it}^{pk}. However, we can make guesses at how the estimated distribution of unobserved heterogeneity in tastes and the price coefficient might change. Jiang's (2013) θ_i parameter, which determines unobserved heterogeneity in taste for calling, functions similarly to our μ_i^{pk}. Our estimates of Jiang's (2013) model in Table 20 imply an average θ_i of 183 minutes, while Jiang's (2013) imply an average θ_i of 202 minutes. Since the average tastes are about the same, we do not think we would get substantially different estimates of the average μ_i^{pk} were we to apply our model to her data. The estimated standard deviation of θ_i from our implementation of Jiang's (2013) model is about 80 minutes, while Jiang's (2013) estimate is 215 minutes. This suggests we would get a larger estimate of the variance of μ_i^{pk} were we to apply our model to her data. If we underestimate the standard deviation of μ_i^{pk} by the same proportion as we underestimate the standard deviation of θ_i, we would want to increase the standard deviation of μ_i^{pk} by about 2.5 times.

Our estimates in Table 20 also imply that we underestimate consumer price sensitivity. One way to get a sense of how much we underestimate price sensitivity is to consider how much a consumer who observes a price increase from 0 to 11 cents on a two part tariff would reduce her usage for each set of estimates. In Jiang's (2013) model, a consumer's target usage on a two part tariff has a closed form solution:

$$x^* = \frac{\theta_i}{a_i p + 1}$$

Our estimates of Jiang's (2013) model imply an average consumer would target usage at 183 minutes if the marginal price were 0, and 92 minutes if the marginal price were 11 cents - a 50 percent reduction in usage. Jiang's (2013) estimates imply an average consumer would target 202 minutes at a price of 0, and 64 minutes at a marginal price of 11 cents - a 68 percent reduction in usage. If our model underestimates the actual reduction in usage by 18 percent, then the price coefficient would have to be a little over 4. This analysis motivates our choice of robustness checks in Appendix G.4 Table 19.

References

Borenstein, Severin, "To What Electricity Price Do Consumers Respond? Residential Demand Elasticity Under Increasing-Block Pricing," 2009. http://faculty.haas.berkeley.edu/borenste/download/NBER_SI_2009.pdf.

DeGroot, Morris H., *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970.

Dutang, Christophe and Petr Savicky, "randtoolbox: Generating and Testing Random Numbers.," 2010. R package version 1.1. http://cran.r-project.org/web/packages/randtoolbox/.

Goettler, Ronald L. and Ron Shachar, "Spatial Competition in the Network Television Industry," *The RAND Journal of Economics*, Winter 2001, *32* (4), 624–656.

Gourieroux, Christian and Alain Monfort, "Simulation-based inference : A survey with special reference to panel data models," *Journal of Econometrics*, 1993, *59* (1-2), 5–33.

Hajivassiliou, Vassilis A. and Paul A. Ruud, "Classical Estimation Methods for LDV Models Using Simulation," in R.F. Engle and Daniel L. McFadden, eds., *Handbook of Econometrics, Volume IV*, Amsterdam: North-Holland, 1994.

Jiang, Lai, "The Welfare Effects of 'Bill Shock' Regulation in Mobile Telecommunication Markets," November 2013. http://www.lai-jiang.com/uploads/Jiang_JobMarketPaper.pdf.

Roodman, David, "How to do xtabond2: An introduction to difference and system GMM in Stata," *Stata Journal*, 2009, *9* (1), 86–136.

TelcoData.us, "npa-nxx-companytype-ocn.csv," 2005. Telecommunications Database. https://www.telcodata.us/data-downloads (accessed October 13, 2005).

Train, Kenneth E., *Discrete Choice Methods with Simulation*, 2nd ed., New York: Cambridge University Press, 2009.