Try It, You'll Like It-Or Will You? The Perils of Early Free-Trial Promotions for High-Tech Service Adoption

Web Appendix

Web Appendix 1: Derivation of $\mathbf{q_{it}^F}$, $\mathbf{q_{it}^R}$, and $\mathbf{UR_{it}}$

Derivation of q_{it}^F and q_{it}^R. q_{it}^F and q_{it}^R represent the beliefs with regard to the average match quality over the months of the free trial and the regular contract, respectively. Below, we derive the distribution of the belief q_{it}^L with regard to the average match quality during a contract of *any* length L. First, notice that:

$$q_{it}^L = \sum_{\tau=1}^{L} q_{it}^{t+\tau}/L = \sum_{\tau=1}^{L}\big(q_{it} + k \cdot \max(0, T - t - \tau)\big)/L =$$

$$q_{it} + k \cdot \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L.$$

Because $q_{it}|d, k \sim N(\bar{q}_{it}, s_{it})$ (see Equations 11 and 12), we can write:

(A.1) $$E\big(q_{it}^L\big|d, k\big) = \bar{q}_{it} + k \cdot \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L \text{ and}$$

(A.2) $$\mathrm{Var}\big(q_{it}^L\big|d, k\big) = s_{it}^2.$$

On the basis of Equation 11, $E\big(q_{it}^L\big|d, k\big)$ can be rewritten as:

(A.3) $$E\big(q_{it}^L\big|d, k\big) = s_{it}^2 \cdot \left[\frac{E(q_0|d)}{\mathrm{Var}(q_0|d)} + \frac{\sum_{\tau=1}^{t} \vartheta_{i\tau}^A}{(\sigma^A)^2} + \frac{\sum_{\tau=1}^{t} \vartheta_{i\tau}^D}{(\sigma^D)^2} + \frac{\sum_{\tau=1}^{t} \vartheta_{i\tau}^U}{(\sigma^U)^2} + \frac{\sum_{\tau=1}^{t} \vartheta_{i\tau}^W}{(\sigma^W)^2}\right] +$$

$$k \cdot \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L,$$

where $q_0|d$ is the initial belief about the ultimate quality θ_i, conditional on d. Consumers derive q_0 as $q_0 = q_0^A - d$, where q_0^A is consumers' initial belief about advertised quality. Indeed, prior to the launch of the service, consumers only receive quality signals through advertising or direct marketing – obviously, there are no usage or WOM signals before launch – such that any knowledge about the ultimate quality θ_i is based on what they have learned from marketing (q_0^A) and their belief with regard to the marketing bias (d). If we assume that $q_0^A \sim N\big(\bar{q}_0^A, s_0\big)$, it follows that $q_0|d \sim N\big(\bar{q}_0^A - d, s_0\big)$. Given Expressions 10, we can now rewrite Equation A.3 as follows:

$$\text{(A.4) } E\left(q_{it}^{L} \middle| d,k\right) = s_{it}^{2} \cdot \left[\frac{\bar{q}_0^{A} - d}{s_0^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{A} + n_{i\tau}^{A}\cdot(\delta - d)\right)}{\left(\sigma^{A}\right)^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{D} + n_{i\tau}^{D}\cdot(\delta - d)\right)}{(\sigma^{D})^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{U} - n_{i\tau}^{U}\cdot k\cdot \max(0,T-\tau)\right)}{(\sigma^{U})^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{W} - n_{i\tau}^{W}\cdot k\cdot \max(0,T-\tau)\right)}{(\sigma^{W})^2} \right] + k \cdot \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L,$$

where $\tilde{\vartheta}_{i\tau}^{A} \sim N\left(n_{i\tau}^{A} \cdot \theta_i, \sqrt{n_{i\tau}^{A}} \cdot \sigma^{A}\right)$, $\tilde{\vartheta}_{i\tau}^{D} \sim N\left(n_{i\tau}^{D} \cdot \theta_i, \sqrt{n_{i\tau}^{D}} \cdot \sigma^{D}\right)$, $\tilde{\vartheta}_{i\tau}^{U} \sim N\left(n_{i\tau}^{U} \cdot (\theta_i + \kappa \cdot \text{intervent}_\tau), \sqrt{n_{i\tau}^{U}} \cdot \sigma^{U}\right)$, and $\tilde{\vartheta}_{i\tau}^{W} \sim N\left(n_{i\tau}^{W} \cdot (\theta_i + \kappa \cdot \text{intervent}_\tau), \sqrt{n_{i\tau}^{W}} \cdot \sigma^{W}\right)$. Thus, $\tilde{\vartheta}_{i\tau}^{A}$, $\tilde{\vartheta}_{i\tau}^{D}$, $\tilde{\vartheta}_{i\tau}^{U}$, and $\tilde{\vartheta}_{i\tau}^{W}$ do not depend on d and k.

Since $q_{it}^{L}|d,k$ is normally distributed and $E\left(q_{it}^{L}|d,k\right)$ is a linear function of d and k, the unconditional distribution of q_{it}^{L} is normal too. We use Equation A.4 to derive the unconditional mean $E\left(q_{it}^{L}\right)$ and variance $\text{Var}\left(q_{it}^{L}\right)$. Because by definition $E\left(q_{it}^{L}\right) = \int E\left(q_{it}^{L}|d,k\right) f(d) f(k) \, dd \, dk$ and since $E\left(q_{it}^{L}|d,k\right)$ is linear in d and k (see Equation A.4), the unconditional mean of q_{it}^{L} can be found by replacing d and k in Equation A.4 with $\bar{d}$ and $\bar{k}$:

$$\text{(A.5) } E\left(q_{it}^{L}\right) \equiv \bar{q}_{it}^{L} = E\left(q_{it}^{L} \middle| d = \bar{d}, k = \bar{k}\right) = s_{it}^{2} \cdot \left[\frac{\bar{q}_0^{A} - \bar{d}}{s_0^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{A} + n_{i\tau}^{A}\cdot(\delta - \bar{d})\right)}{\left(\sigma^{A}\right)^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{D} + n_{i\tau}^{D}\cdot(\delta - \bar{d})\right)}{(\sigma^{D})^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{U} - n_{i\tau}^{U}\cdot \bar{k}\cdot \max(0,T-\tau)\right)}{(\sigma^{U})^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{W} - n_{i\tau}^{W}\cdot \bar{k}\cdot \max(0,T-\tau)\right)}{(\sigma^{W})^2} \right] + \bar{k} \cdot \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L.$$

As becomes clear from Equation A.5, $\bar{q}_0^{A}$, $\bar{d}$, and δ cannot be separately identified. Therefore we estimate $\bar{q}_0$ and $\bar{d}^{*}$ with $\bar{q}_0 = \bar{q}_0^{A} - \bar{d}$ (i.e., the unconditional initial mean quality belief about θ_i) and $\bar{d}^{*} = \delta - \bar{d}$.

Next, we derive $\mathrm{Var}\left(q_{it}^{L}\right)$, that is, the consumers' uncertainty at time t – given the signals received by that time – about the average monthly quality over the next L months. To find $\mathrm{Var}\left(q_{it}^{L}\right)$, notice that we can write Equation A.4 as:

$$\text{(A.6)} \qquad E\left(q_{it}^{L}\middle|d,k\right) = F_{it} + G_{it} \cdot (\delta - d) + H_{it} \cdot k,$$

with $F_{it} = s_{it}^{2} \cdot \left(\frac{\bar{q}_{0}^{A} - \delta}{s_{0}^{2}} + \frac{\sum_{\tau=1}^{t} \tilde{\vartheta}_{i\tau}^{A}}{\left(\sigma^{A}\right)^{2}} + \frac{\sum_{\tau=1}^{t} \tilde{\vartheta}_{i\tau}^{D}}{(\sigma^{D})^{2}} + \frac{\sum_{\tau=1}^{t} \tilde{\vartheta}_{i\tau}^{U}}{(\sigma^{U})^{2}} + \frac{\sum_{\tau=1}^{t} \tilde{\vartheta}_{i\tau}^{W}}{(\sigma^{W})^{2}}\right)$, $G_{it} = s_{it}^{2} \cdot \left(\frac{1}{s_{0}^{2}} + \frac{\sum_{\tau=1}^{t} n_{i\tau}^{A}}{\left(\sigma^{A}\right)^{2}} + \frac{\sum_{\tau=1}^{t} n_{i\tau}^{D}}{(\sigma^{D})^{2}}\right)$, and $H_{it} = \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L - s_{it}^{2} \cdot \frac{\sum_{\tau=1}^{t} n_{i\tau}^{U} \cdot \max(0, T-\tau)}{(\sigma^{U})^{2}} - s_{it}^{2} \cdot \frac{\sum_{\tau=1}^{t} n_{i\tau}^{W} \cdot \max(0, T-\tau)}{(\sigma^{W})^{2}}$.

Using the law of total variance Var(X) = E(Var(X|Y)) + Var(E(X|Y)) (Weiss, 2005), we can now derive $\mathrm{Var}\left(q_{it}^{L}\right)$ as follows:

$$\text{(A.7)} \quad \mathrm{Var}\left(q_{it}^{L}\right) \equiv \left(s_{it}^{L}\right)^{2} = E\left(s_{it}^{2}\right) + \mathrm{Var}\left(F_{it} + G_{it} \cdot (\delta - d) + H_{it} \cdot k\right)$$

$$= s_{it}^{2} + G_{it}^{2} \cdot \left(s^{d}\right)^{2} + H_{it}^{2} \cdot \left(s^{k}\right)^{2},$$

where the last equality is based on the fact that s_{it}^{2} does not depend on d and k, $\mathrm{Var}(\delta - d) = \left(s^{d}\right)^{2}$, and $\mathrm{Var}(k) = \left(s^{k}\right)^{2}$.

Derivation of UR_{it}. Consider a consumer who, at time t, has to decide on whether or not to accept a three-month free trial offer. Let L^{R} be the length of the contractual period if he adopts the service after the trial, and let $\left(s_{i(t+3)}^{R}\right)^{2}$ be the consumer's uncertainty at time t + 3 about the average match quality during the contractual period of L^{R} months, starting in month t + 4.

To assess the likely reduction in $\left(s_{i(t+3)}^{R}\right)^{2}$ as a result of the expected usage intensity during the trial, the consumer needs to assess $\left(s_{i(t+3)}^{R}\right)^{2}$ with and without usage signals, and compute the difference. Notice that, in line with Equation A.7:

$$\text{(A.8)} \qquad \left(s_{i(t+3)}^{R}\right)^{2} = s_{i(t+3)}^{2} + G_{i(t+3)}^{2} \cdot \left(s^{d}\right)^{2} + H_{i(t+3)}^{2} \cdot \left(s^{k}\right)^{2},$$

with $G_{i(t+3)} = s^2_{i(t+3)} \cdot \left(\frac{1}{s_0^2} + \frac{\sum_{\tau=1}^{t+3} n^A_{i\tau}}{(\sigma^A)^2} + \frac{\sum_{\tau=1}^{t+3} n^D_{i\tau}}{(\sigma^D)^2}\right)$, $H_{i(t+3)} = \sum_{\tau=1}^{L^R} \max(0, T-(t+3+\tau))/L^R - s^2_{i(t+3)} \cdot \frac{\sum_{\tau=1}^{t+3} n^U_{i\tau} \cdot \max(0, T-\tau)}{(\sigma^U)^2} - s^2_{i(t+3)} \cdot \frac{\sum_{\tau=1}^{t+3} n^W_{i\tau} \cdot \max(0, T-\tau)}{(\sigma^W)^2}$. Note that a consumer can only *predict* $\left(s^R_{i(t+3)}\right)^2$ because, in month t, she does not know the true values of $n^A_{i(t+k)}$, $n^D_{i(t+k)}$, $n^U_{i(t+k)}$, and $n^W_{i(t+k)}$ for k = 1, 2, 3. In our model, the consumer predicts $n^A_{i(t+k)}$ and $n^D_{i(t+k)}$ using the average monthly number of past advertising and direct marketing signals, $\sum_{\tau=1}^{t} n^A_{i\tau}/t$ and $\sum_{\tau=1}^{t} n^D_{i\tau}/t$, repectively. For $n^U_{i(t+k)}$, we assume that the consumer has a notion of her likely monthly use rate. We compute this likely use rate as the consumer's average monthly usage across her subscription months. For those consumers who never subscribed to the service, we predict the likely use rate with a loglinear regression of zaps on age, average income, household size, and relationship length, calibrated on the data of actual IDTV users. Of course, predicting $\mathrm{Var}\left(q^R_{i(t+3)}\right)$ *in the absence of usage signals* is a matter of setting $n^U_{i(t+k)}$ to zero. For $n^W_{i(t+k)}$, the consumer first computes the average monthly increase in WOM, $\overline{\Delta n}^W_{it} = \left(n^W_{it} - n^W_{i1}\right)/(t-1)$, and then predicts $n^W_{i(t+k)}$ as $n^W_{it} + k \cdot \overline{\Delta n}^W_{it}$. Finally, to obtain UR_{it}, we center each customer's series of uncertainty reduction values around the value in the first period, such that UR_{it} only captures within-consumer cross-time variation in anticipated uncertainty reduction.

Web Appendix 2: Robustness of the Reduced Form Approach

The reduced-form approach has important advantages over the structural approach. For one, it uses a closed-form expression for the "expected value function" or the "gain from trial," which is easy to interpret and speeds up estimation time (Ching et al., 2011). Moreover, it avoids the identification difficulties that typically come with a fully structural approach (Ching et al., 2013), especially if the value outcomes of consumers' decisions are not directly observed (as is the case here: we do not observe the utility that consumers derive from service adoption). However, our reduced form approach also comes with limitations. In reality, the expected gain from trial may be more complex than only a linear function of the expected uncertainty reduction (see Equation 1). Specifically, our model ignores any interactions between the expected uncertainty reduction and the *level* of the quality beliefs (Ching et al., 2011). The estimated coefficient of our reduced-form uncertainty reduction (which is constant for a given consumer) does not capture this, and thus represents some average effect across observations. To assess the accuracy and robustness of the reduced form, we run a simulation in which we compare the anticipated uncertainty reduction due to trial, UR_{it} (see Web Appendix 1), with the anticipated benefit of trial as obtained in a structural forward-looking approach. We first describe the structural model and then discuss the simulation results.

Structural model. We consider a consumer at time t, who has not adopted the service yet, but has the opportunity to benefit from a trial offer. She knows that accepting the trial will also have consequences beyond the trial period, in that it will allow her to make a more informed decision on whether or not to fully adopt at time t + 3 (after the trial expired). In our reduced form approach, this is captured by including the anticipated uncertainty reduction UR_{it}, i.e., the drop in uncertainty about the average match quality during a subsequent paid contractual period, due to trial usage.

In the structural approach, the consumer evaluates the extent to which subscribing to the free trial affects the expected utility three months from now. If $\widehat{V}_{i,t+3}^{R(\text{After Trial})}$ refers to the consumer's estimated systematic utility from adopting the regular paid service after first having used the trial (see Equation 2), and given that the utility of not accepting the paid offer is always scaled to zero, the expected maximum utility in t + 3 can be found as $\ln\left(\exp\left(\widehat{V}_{i,t+3}^{R(\text{After Trial})}\right) +\right.$

$1\Big)$ (Train, 2009). Similarly, $\ln\left(\exp\left(\hat{V}_{i,t+3}^{R(\text{Without Trial})}\right)+1\right)$ is the expected maximum utility in t + 3 when the consumer did not first use a trial. The future gain from trial is then obtained as the difference between these two expressions, which we henceforth refer to as DV_{it}.

The challenge is to correctly derive $\hat{V}_{i,t+3}^{R(\text{After Trial})}$ and $\hat{V}_{i,t+3}^{R(\text{Without Trial})}$. In line with Equations 2 and 4, we write:

(A.9) $$\hat{V}_{i,t+3}^{R(\text{After Trial})}=\beta_i^0-\exp\left[-r\left(\hat{q}_{i(t+3)}^{R(\text{After Trial})}-r\left(\hat{s}_{i(t+3)}^{R(\text{After Trial})}\right)^2/2\right)\right]+\beta_i^{FE}+\hat{X}_{i(t+3)}\beta_i^X$$

and

(A.10) $$\hat{V}_{i,t+3}^{R(\text{Without Trial})}=\beta_i^0-\exp\left[-r\left(\hat{q}_{i(t+3)}^{R(\text{Without Trial})}-r\left(\hat{s}_{i(t+3)}^{R(\text{Without Trial})}\right)^2/2\right)\right],$$

$$+\beta_i^{DISC}\widehat{DISC}_{i(t+3)}+\hat{X}_{i(t+3)}\beta_i^X.$$

where $\hat{q}_{i,t+3}^{R(\text{After Trial})}$ and $\hat{s}_{i,t+3}^{R(\text{After Trial})}$ are the anticipated mean and standard deviation of the consumer's belief regarding the service quality during a regular contract, after first having used a three-month trial. Similarly, $\hat{q}_{i,t+3}^{R(\text{Without Trial})}$ and $\hat{s}_{i,t+3}^{R(\text{Without Trial})}$ are the anticipated mean and standard deviation of her belief regarding the service quality during a regular contract starting in three months, *without* first having subscribed to a free trial. Notice that we assume the consumer knows the coefficients of the utility functions. Also, she is aware of the fact that, after a trial, she is excluded from any discounts. However, for all other elements of the utility functions, the consumer is relying on estimates. This is particularly challenging for $\hat{q}_{i,t+3}^{R(\text{After Trial})}$, $\hat{q}_{i,t+3}^{R(\text{Without Trial})}$, $\hat{s}_{i,t+3}^{R(\text{After Trial})}$, and $\hat{s}_{i,t+3}^{R(\text{Without Trial})}$, which are the outcome of an anticipated learning process. In Web Appendix 1, we already explained how $\left(\hat{s}_{i,t+3}^{R(\text{After Trial})}\right)^2$ and $\left(\hat{s}_{i,t+3}^{R(\text{Without Trial})}\right)^2$ can be derived (they form the basis for the computation of UR_{it}). For the derivation of $\hat{q}_{i,t+3}^{R(\text{After Trial})}$ and $\hat{q}_{i,t+3}^{R(\text{Without Trial})}$, we rely on Equation A.5 (see Web Appendix 1). Specifically, notice that:

$$\text{(A.11)} \qquad \hat{q}_{i,t+3}^{R(\text{After Trial})} = \left(\hat{\mathbf{s}}_{\mathbf{i,t+3}}^{\textbf{After Trial}}\right)^{\mathbf{2}} \cdot \left[\frac{\bar{q}_0^A - \bar{d}}{s_0^2} + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{A} + n_{i\tau}^{A}\cdot(\delta - \bar{d})\right) + \sum_{\tau=1}^{3}\left(\widehat{\tilde{\boldsymbol{\vartheta}}_{\mathbf{i,t+\tau}}^{\mathbf{A}} + \mathbf{n}_{\mathbf{i,t+\tau}}^{\mathbf{A}}}\cdot(\boldsymbol{\delta} - \bar{\mathbf{d}})\right)}{(\sigma^A)^2}\right.$$

$$+ \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{D} + n_{i\tau}^{D}\cdot(\delta - \bar{d})\right) + \sum_{\tau=1}^{3}\left(\widehat{\tilde{\boldsymbol{\vartheta}}_{\mathbf{i,t+\tau}}^{\mathbf{D}} + \mathbf{n}_{\mathbf{i,t+\tau}}^{\mathbf{D}}}\cdot(\boldsymbol{\delta} - \bar{\mathbf{d}})\right)}{(\sigma^D)^2}$$

$$+ \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{U} - n_{i\tau}^{U}\cdot\bar{k}\cdot\max(0, T-\tau)\right) + \sum_{\tau=1}^{3}\left(\tilde{\boldsymbol{\vartheta}}_{\mathbf{i,t+\tau}}^{\mathbf{U}} - \widehat{\mathbf{n}_{\mathbf{i,t+\tau}}^{\mathbf{U}}}\cdot\bar{\mathbf{k}}\cdot\mathbf{max}(\mathbf{0}, \mathbf{T-t-\tau})\right)}{(\sigma^U)^2}$$

$$\left. + \frac{\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{W} - n_{i\tau}^{W}\cdot\bar{k}\cdot\max(0, T-\tau)\right) + \sum_{\tau=1}^{3}\left(\tilde{\boldsymbol{\vartheta}}_{\mathbf{i,t+\tau}}^{\mathbf{W}} - \widehat{\mathbf{n}_{\mathbf{i,t+\tau}}^{\mathbf{W}}}\cdot\bar{\mathbf{k}}\cdot\mathbf{max}(\mathbf{0}, \mathbf{T-t-\tau})\right)}{(\sigma^W)^2}\right]$$

$$+ \bar{k} \cdot \sum_{\tau=1}^{R} \max(0, T - t - 3 - \tau)/R,$$

with

$$\text{(A.12)} \qquad \left(\hat{s}_{i(t+3)}^{\text{After Trial}}\right)^2 = \left(\frac{1}{s_{it}^2} + \frac{\sum_{\tau=1}^{3} \hat{\mathbf{n}}_{\mathbf{i(t+\tau)}}^{\mathbf{A}}}{(\sigma^A)^2} + \frac{\sum_{\tau=1}^{3} \hat{\mathbf{n}}_{\mathbf{i(t+\tau)}}^{\mathbf{D}}}{(\sigma^D)^2} + \frac{\sum_{\tau=1}^{3} \hat{\mathbf{n}}_{\mathbf{i(t+\tau)}}^{\mathbf{U}}}{(\sigma^U)^2} + \frac{\sum_{\tau=1}^{3} \hat{\mathbf{n}}_{\mathbf{i(t+\tau)}}^{\mathbf{W}}}{(\sigma^W)^2}\right)^{-1}.$$

The expressions for $\hat{q}_{i,t+3}^{R(\text{Without Trial})}$ and $\left(\hat{s}_{i,t+3}^{\text{Without Trial}}\right)^2$ are analogous, except that they do not account for learning from usage signals.[1] Notice that the components in bold can only be *estimated* since they involve information that is unknown to the consumer in month t. For the derivation of $\hat{s}_{i,t+3}^{\text{After Trial}}$, she only needs to estimate the numbers of signals $n_{i,t+\tau}^{A}$, $n_{i,t+\tau}^{D}$, $n_{i,t+\tau}^{U}$, and $n_{i,t+\tau}^{W}$ during the next three months, e.g., on the basis of her signal history or consumer characteristics (see Web Appendix 1). The derivation of the components in bold between the

---

[1] In the third term of the expression for $\hat{q}_{i,t+3}^{R(\text{After Trial})}$ we can in principle leave out $\sum_{\tau=1}^{t}\left(\tilde{\vartheta}_{i\tau}^{U} - n_{i\tau}^{U} \cdot \bar{k} \cdot \max(0, T - \tau)\right)$ because, in the current analysis, we focus on consumers who consider to accept or not accept a trial in t without having used the service before and, thus, who have not received any usage signals before t + 1.

square brackets is more complex because not only is the consumer unaware of the exact values of the quality signals during the next three months, she does not even know the distributions that generate these signals. Thus, the consumer cannot use the real signal-generating distributions to integrate out the quality signals.

Let us first discuss $\sum_{\tau=1}^{3}\left(\widehat{\tilde{\vartheta}_{i,t+\tau}^{A} + n_{i,t+\tau}^{A}} \cdot (\delta - \bar{d})\right)$ and $\sum_{\tau=1}^{3}\left(\widehat{\tilde{\vartheta}_{i,t+\tau}^{D} + n_{i,t+\tau}^{D}} \cdot (\delta - \bar{d})\right)$, capturing the impact of the corrected advertising and direct marketing signals in months t + 1, t + 2, and t + 3. From Web Appendix 1, we remember that, in principle, $\tilde{\vartheta}_{i,t+\tau}^{A} \sim N\left(n_{i,t+\tau}^{A} \cdot \theta_i, \sqrt{n_{i,t+\tau}^{A}} \cdot \sigma^{A}\right)$ and $\tilde{\vartheta}_{i,t+\tau}^{D} \sim N\left(n_{i,t+\tau}^{D} \cdot \theta_i, \sqrt{n_{i,t+\tau}^{D}} \cdot \sigma^{D}\right)$, but the consumer does not know $n_{i,t+\tau}^{A}$, $n_{i,t+\tau}^{D}$, or her real ultimate match quality θ_i. In our structural model, she therefore replaces $\tilde{\vartheta}_{i,t+\tau}^{A}$ and $\tilde{\vartheta}_{i,t+\tau}^{D}$ with $\hat{\vartheta}_{i,t+\tau}^{A}$ and $\hat{\vartheta}_{i,t+\tau}^{D}$, where $\hat{\vartheta}_{i,t+\tau}^{A} \big| q_{it} \sim N\left(\hat{n}_{i,t+\tau}^{A} \cdot q_{it}, \sqrt{\hat{n}_{i,t+\tau}^{A}} \cdot \sigma^{A}\right)$ and $\hat{\vartheta}_{i,t+\tau}^{D} \big| q_{it} \sim N\left(\hat{n}_{i,t+\tau}^{D} \cdot q_{it}, \sqrt{\hat{n}_{i,t+\tau}^{D}} \cdot \sigma^{D}\right)$, and q_{it} is the consumer's current belief with regard to the ultimate match quality. The consumer's assessment of $\sum_{\tau=1}^{3}\left(\widehat{\tilde{\vartheta}_{i,t+\tau}^{A} + n_{i,t+\tau}^{A}} \cdot (\delta - \bar{d})\right)$ and $\sum_{\tau=1}^{3}\left(\widehat{\tilde{\vartheta}_{i,t+\tau}^{D} + n_{i,t+\tau}^{D}} \cdot (\delta - \bar{d})\right)$ is further complicated by the fact that she does not know δ, the real marketing bias. Instead, she therefore relies on her bias belief $d \sim N(\bar{d}, s^{d})$. In summary:

(A.13) $$\sum_{\tau=1}^{3}\left(\widehat{\tilde{\vartheta}_{i(t+\tau)}^{A} + n_{i(t+\tau)}^{A}} \cdot (\delta - \bar{d})\right) \equiv \sum_{\tau=1}^{3}\left(\hat{\vartheta}_{i(t+\tau)}^{A} + \hat{n}_{i(t+\tau)}^{A} \cdot (d - \bar{d})\right) \text{ and}$$

(A.14) $$\sum_{\tau=1}^{3}\left(\widehat{\tilde{\vartheta}_{i(t+\tau)}^{D} + n_{i(t+\tau)}^{D}} \cdot (\delta - \bar{d})\right) \equiv \sum_{\tau=1}^{3}\left(\hat{\vartheta}_{i(t+\tau)}^{D} + \hat{n}_{i(t+\tau)}^{D} \cdot (d - \bar{d})\right).$$

Similar observations pertain to the expressions

$\sum_{\tau=1}^{3}\left(\widetilde{\vartheta}_{\imath,t+\tau}^{U}-\widehat{n_{\imath,t+\tau}^{U}\cdot\bar{k}}\cdot\max(0,T-t-\tau)\right)$ and $\sum_{\tau=1}^{3}\left(\widetilde{\vartheta}_{\imath,t+\tau}^{W}-\widehat{n_{\imath,t+\tau}^{W}\cdot\bar{k}}\cdot\max(0,T-t-\tau)\right)$, capturing the impact of the corrected usage and WOM signals in months t + 1, t + 2, and t + 3. According to our derivations in Web Appendix 1, it should hold that $\widetilde{\vartheta}_{i,t+\tau}^{U}\sim N\left(n_{i,t+\tau}^{U}\cdot(\theta_i+\kappa\cdot\mathrm{intervent}_{t+\tau}),\sqrt{n_{i,t+\tau}^{U}}\cdot\sigma^{U}\right)$ and $\widetilde{\vartheta}_{i,t+\tau}^{W}\sim N\left(n_{i,t+\tau}^{W}\cdot(\theta_i+\kappa\cdot\mathrm{intervent}_{t+\tau}),\sqrt{n_{i,t+\tau}^{W}}\cdot\sigma^{W}\right)$. However, the consumer does not know the numbers of signals $n_{i,t+\tau}^{U}$ and $n_{i,t+\tau}^{W}$, the true ultimate match quality θ_i, and $\kappa\cdot\mathrm{intervent}_{t+\tau}$, i.e., the extent to which the current match quality differs from the ultimate match quality. Therefore, she replaces $\widetilde{\vartheta}_{i,t+\tau}^{U}$ and $\widetilde{\vartheta}_{i,t+\tau}^{W}$ with $\widehat{\vartheta}_{i,t+\tau}^{U}$ and $\widehat{\vartheta}_{i,t+\tau}^{W}$, where $\widehat{\vartheta}_{i,t+\tau}^{U}\Big|\,q_{it},k\sim N\left(\hat{n}_{i,t+\tau}^{U}\cdot\left(q_{it}+k\cdot\max(0,T-t-\tau)\right),\sqrt{\hat{n}_{i,t+\tau}^{U}}\cdot\sigma^{U}\right)$, $\widehat{\vartheta}_{i,t+\tau}^{W}\Big|\,q_{it},k\sim N\left(\hat{n}_{i,t+\tau}^{W}\cdot\left(q_{it}+k\cdot\max(0,T-t-\tau)\right),\sqrt{\hat{n}_{i,t+\tau}^{W}}\cdot\sigma^{W}\right)$, q_{it}, like before, is the consumer's current belief with regard to the ultimate match quality, and $k\sim N\left(\bar{k},s^{k}\right)$ is the belief with regard to the monthly quality change. In summary, we have:

(A.15) $\sum_{\tau=1}^{3}\left(\widetilde{\vartheta}_{\imath(t+\tau)}^{U}-\widehat{n_{\imath(t+\tau)}^{U}\cdot\bar{k}}\cdot\max(0,T-t-\tau)\right)\equiv\sum_{\tau=1}^{3}\left(\widehat{\vartheta}_{i(t+\tau)}^{U}-\hat{n}_{i(t+\tau)}^{U}\cdot\bar{k}\cdot\max(0,T-t-\tau)\right)$

and

(A.16) $\sum_{\tau=1}^{3}\left(\widetilde{\vartheta}_{\imath,t+\tau}^{W}-\widehat{n_{\imath,t+\tau}^{W}\cdot\bar{k}}\cdot\max(0,T-t-\tau)\right)\equiv\sum_{\tau=1}^{3}\left(\widehat{\vartheta}_{i,t+\tau}^{W}-\hat{n}_{i,t+\tau}^{W}\cdot\bar{k}\cdot\max(0,T-t-\tau)\right)$.

In the structural model, the consumer derives the expected utility gain from trial by integrating out $\sum_{\tau=1}^{3}\left(\hat{\vartheta}_{i,t+\tau}^{A}+\hat{n}_{i,t+\tau}^{A}\cdot\left(d-\bar{d}\right)\right)$, $\sum_{\tau=1}^{3}\left(\hat{\vartheta}_{i,t+\tau}^{D}+\hat{n}_{i,t+\tau}^{D}\cdot\left(d-\bar{d}\right)\right)$, $\sum_{\tau=1}^{3}\left(\hat{\vartheta}_{i,t+\tau}^{U}-\hat{n}_{i,t+\tau}^{U}\cdot\bar{k}\cdot\max(0,T-t-\tau)\right)$, and $\sum_{\tau=1}^{3}\left(\hat{\vartheta}_{i,t+\tau}^{W}-\hat{n}_{i,t+\tau}^{W}\cdot\bar{k}\cdot\max(0,T-t-\tau)\right)$ from $DV_{it}=$

$\ln\left(\exp\left(\widehat{V}_{i,t+3}^{R(\text{After Trial})}\right)+1\right)-\ln\left(\exp\left(\widehat{V}_{i,t+3}^{R(\text{Without Trial})}\right)+1\right)$. Relying on the principles that we also used in Web Appendix 1, we can derive the unconditional distributions for each of those four expressions. Because the derivations, though similar in nature to those in Web Appendix 1, are lengthy and tedious, we do not completely report them here. In our computation of DV_{it}, we draw from the unconditional distributions to integrate out those four quantities.

Simulation. To assess the robustness of UR_{it}, we examine its cross-time correlation with DV_{it}. Specifically, we consider a consumer with mean values for the (heterogeneous) parameters and socio-demographic, discount, usage, advertising, and direct marketing variables, and month-specific average values for the WOM measure. For this "average consumer," we assess UR_{it} and DV_{it} over the simulation horizon and find a correlation of .94. If we only consider the months in which the trial was actually available, the correlation even further increases to .99. To assess the robustness of UR_{it} against changes in usage, we vary usage intensity from the bottom tenth percentile to the top tenth percentile by increments of 10% and find the correlation between UR_{it} and DV_{it} to remain virtually unaffected. Hence, the reduced-form expression for the forward-looking component (UR_{it}) provides a good approximation of the structural "gain from trial," even for trial timings that are not actually observed, and for different levels of use rate. This instills confidence in the approximation, and suggests that it can be safely used for policy simulations. To further check the sensitivity of our results to UR_{it}, we rerun the policy simulations while systematically increasing or decreasing the impact of UR_{it} in Equation 1 by 20%. Although this influences the amplitude of the simulated effects, it leaves our insights regarding the role of timing and usage unaffected.

Web Appendix 3: Identification of Model Parameters

We first discuss the identification of the exponential Constant Absolute Risk Aversion (CARA) specifications (see Equations 3 and 4) and the underlying learning process, and then deal with the remaining coefficients in Equations 1 and 2.

Identification of the parameters of the CARA specification and underlying learning process. The identification of the distribution parameters of r_i is aided by the fact that, in the CARA specification, r_i interacts with the variance of the consumer's quality belief $\left(s_{it}^{L}\right)^2$. In contrast with the *mean* of the consumer's quality belief $\overline{q}_{it}^{L}$ (which fluctuates with the random signal draws), the belief variance declines monotonically. Because r_i interacts with $\left(s_{it}^{L}\right)^2$, a decrease in variance will result in a stronger utility increase for customers with a high risk-aversion parameter than for less risk-averse customers. This dependence on the evolution of the belief variance sets (the heterogeneity in) the risk aversion parameter apart from (the heterogeneity in) the intercept and the ultimate match quality (which will be discussed below). Furthermore, notice that all consumers are assumed to start with the same quality belief (mean and variance) such that differences in behavior in the early periods can be explained mainly by differences in risk aversion, all else equal (Narayanan & Manchanda, 2009).

The identification of the distribution parameters of the ultimate match quality θ_i is complicated by two factors. First, consumers do not base their decisions on the belief q_{it} about θ_i as such, but on the belief about the average *actual* quality during the (trial or regular) subscription (thereby taking into account that this actual quality may change over time). This belief about the average actual quality, given by $q_{it}^{L} = q_{it} + k \cdot \sum_{\tau=1}^{L} \max(0, T - t - \tau)/L$, is only a *function* of the belief q_{it} about θ_i. Second, the means of the corrected quality signals ($\theta_i + \bar{d}^*$ for advertising and direct marketing and $\theta_i + \kappa \cdot \mathrm{intervent}_t - \bar{k} \cdot \max(0, T - t)$ for usage and WOM), which are used to update q_{it}, do not necessarily correspond to θ_i: that is, consumers may over- or undercorrect for possible signal bias. To solve the first concern, notice that, as a consumer cumulates signals, her belief q_{it} about θ_i converges and the belief's variance s_{it}^2 tends to zero.[2] When convergence has been reached, changes in the belief q_{it}^L merely depend on the passage of time (captured by $\sum_{\tau=1}^{L} \max(0, T - t - \tau)/L$), which enables us to separate consumers' belief $k \sim N(\bar{k}, s^k)$ from q_{it}.[3]

With regard to the second concern, we point out that our dataset offers sufficient variation to separate the ultimate match quality θ_i from the signal bias (correction) parameters $d^* \sim N(\bar{d}^*, s^d)$, κ, and k. The heterogeneity in θ_i (as captured by its standard deviation σ^θ) can be assessed by comparing consumers who collected the same numbers of signals at the same points in time, such that differences in behavior cannot be due to differences in the information set. Notice that our dataset indeed includes clusters of consumers with very similar signal patterns. For example, 1,317 consumers in our dataset received neither direct marketing nor usage signals (before adoption) and each of them can be matched with (up to 8) consumers living in the same census block such that they received exactly the same numbers of advertising and WOM signals at any given point in time. For consumers with similar signal patterns, the impact of d^*, κ, and k is comparable such that we can attribute differences in quality beliefs to heterogeneity in θ_i. Furthermore, given σ^θ, it is possible to separate $\bar{d}^*$ from the average ultimate match quality $\bar{\theta}$ by comparing households with the same usage and WOM signal patterns but with different total numbers of advertising and/or direct marketing signals; toward the end of our observation period

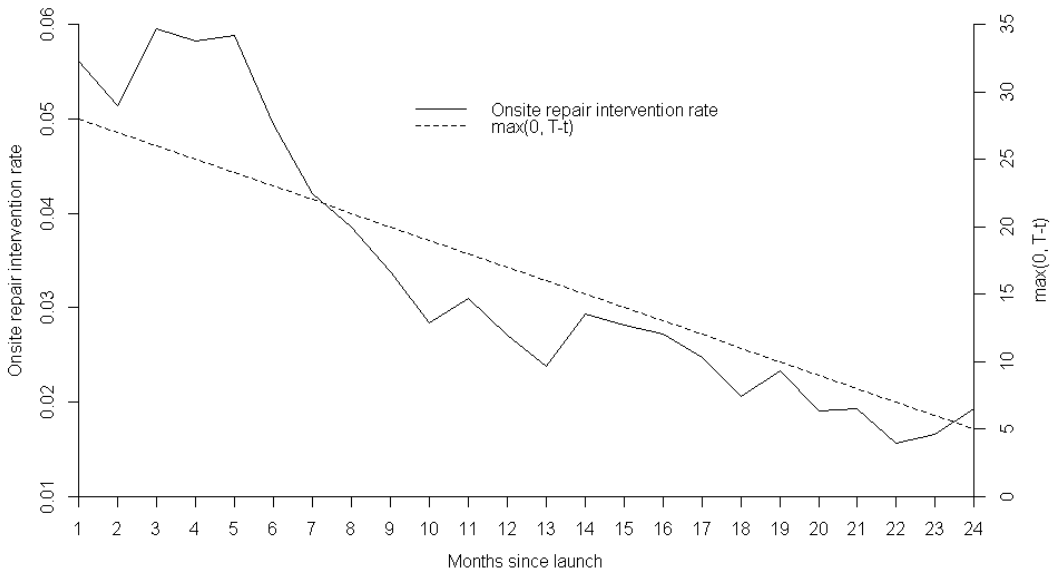
---

[2] Because consumers leave the dataset once they adopt the paid service or opt out after a free trial, there will be consumers for which q_{it} never converges. However, we note that 79% of the consumers do not adopt and 78% of the service subscription observations occur in the second half of our observation period. In other words, there is substantial variability in the dependent variable at a moment that many consumers' quality beliefs q_{it} about ultimate match quality may have converged.

[3] In the CARA specification, expectation and variance operate mathematically independently, such that both $\bar{k}$ and s^k can be identified. Furthermore, for the identification of k, we can rely on the complete consumer sample because k is not heterogeneous.

(when convergence in q_{it} has been reached), the confrontation of cross-sectional differences in subscription behavior with cross-sectional differences in the total number of advertising and direct marketing signals will reveal the average extent $\bar{d}^*$ to which consumers over- or undercorrect for marketing bias. That is, the belief q_{it}^L of consumers with larger total numbers of marketing signals will be affected by $\bar{d}^*$ to a greater degree. For example, in month 19; 5,058 of the consumers who have not adopted yet can be matched with (up to 13) consumers with *identical* (zero-)usage and WOM signal patterns but substantially different total amounts of direct marketing (average within-cluster variance is 16.322). Similarly, we can compare consumers with the same total numbers of advertising and direct marketing signals but with different usage and/or WOM signal patterns. Toward the end of the observation period, such comparison helps us disentangle the impact of κ and $\bar{k}$. More specifically, the belief of consumers who received a large number of usage and WOM signals at some earlier point t, will be affected more by the net signal correction $\kappa \cdot \text{intervent}_t - \bar{k} \cdot \max(0, T - t)$ than the belief of consumers who received fewer signals at the same point in time. For example, in month 19; 1,836 of the consumers who have not adopted yet, have the same advertising and direct marketing history but very different usage and WOM patterns: the average within-period cross-consumer variance is 17,910.253 for the monthly number of WOM signals and 310.125 for the monthly number of usage signals. The parameter κ can be distinguished from $\bar{k}$ because the number of actual interventions (which identifies κ) does not smoothly follow the function $\max(0, T - t)$ (which identifies $\bar{k}$). This becomes clear from Figure A.1 which displays the evolution of the onsite repair intervention rate and $\max(0, T - t)$.

Figure A.1: Data Pattern for Onsite Repair Interventions and $\max(0, T - t)$



Given the other learning parameters, the cross-time subscription behaviors, combined with the substantial variation in the numbers of signals (see illustrations in Figures A.2 and A.3), help us identify the signal standard deviations σ^A, σ^D, σ^U, and σ^W (Coscelli & Shum, 2004). During the first months after launch, the free trial was not available yet, and WOM was limited. In addition, as indicated before, many consumers did not receive any direct marketing. For them, advertising and adoption data from the first months help identify the standard deviation σ^A of the advertising signals. Given σ^A, we can use direct marketing and adoption data for consumers who did receive direct marketing during the first nine months to identify the standard deviation σ^D of the direct marketing signals. Subsequent changes in consumers' adoption as they receive usage and WOM signals help us identify how much they learn from these signals. Since many consumers learned through WOM and *not* usage (almost 85% of the consumers did not subscribe to a trial), we can disentangle the standard deviations σ^U and σ^W of the usage and WOM signals.

Figure A.2: Temporal Variation in Number of Quality Signals

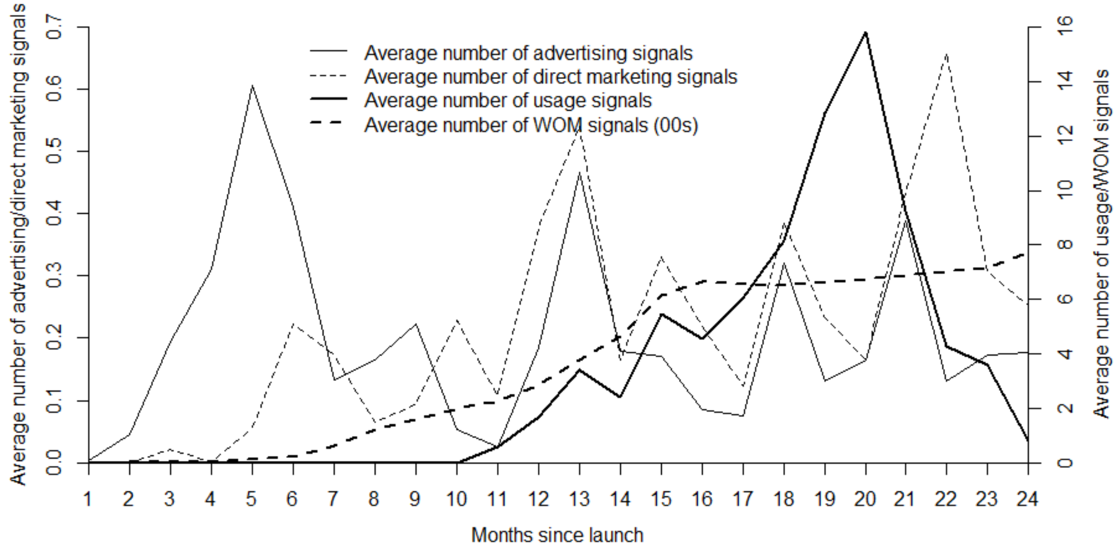
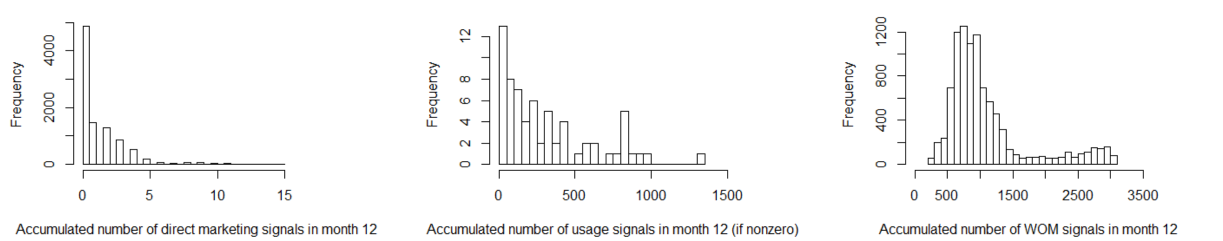


Figure A.3: Cross-sectional Variation in Number of Accumulated Quality Signals in Month 12[4]

a) Direct Marketing b) Usage c) WOM



In a similar way, the combination of consumers' subscription behavior and the numbers of quality signals allows us to identify the standard deviations of the bias beliefs, s^d and s^k. As indicated by Mehta, Chen, & Narasimhan (2008a, 2008b), the standard deviations of the bias

[4] As indicated in Table 1, the cross-sectional variation in advertising is very limited.

beliefs capture the correlations between the corrected signals, and thus have an impact on the signals' informativeness: the higher the bias uncertainty, the higher the variance shared by the corrected signals, and thus the lower the informativeness of a set of corrected signals, relative to a set of completely independent signals. While a comparison of marketing and non-marketing quality sources (e.g., advertising versus WOM) sheds light on the informational value of a single signal (as captured by the signal standard deviations, see above), a comparison of different signal numbers for a *given* signal source reveals the decrease in the marginal informational signal value as the number of signals increases. In other words, a comparison of different signal numbers reveals the correlation between subsequent signals and thus enables us to assess s^d and s^k (see Mehta, Chen, & Narasimhan, 2008a).

Finally, an advantage of our study context is that we start observing all consumers directly after service launch such that we do not face the "left-truncation problem" in which information about initial conditions is lacking. As a result, the observations in the early months are particularly helpful in the identification of some of our parameters. Specifically, in our model, all consumers start with the same initial quality belief $N(\bar{q}_0, s_0)$ about ultimate match quality where s_0 is .01. Shortly after launch, this belief is hardly affected by any quality signals.[5] On the basis of Equations A.5 and A.7, we find that, without quality signals, the consumer's expected service quality $q_{it}^R \sim N\left(\bar{q}_{it}^R, s_{it}^R\right)$ has the following mean and variance:

(A.17) $$\bar{q}_{it}^R = \bar{q}_0 + \bar{k} \cdot \sum_{\tau=1}^{L^R} \max(0, T - t - \tau)/L^R$$

(A.18) $$\left(s_{it}^R\right)^2 = s_0^2 + \left(s^d\right)^2 + \left(\sum_{\tau=1}^{L^R} \max(0, T - t - \tau)/L^R\right)^2 \cdot \left(s^k\right)^2.$$

[5] Initially, the trial was not available yet (such that people could not collect usage signals before adoption), consumers obviously received few WOM signals, there were almost no direct marketing contacts, and the per capita advertising investment was limited and identical for consumers in a given region – in one region, the advertising investment in the first month was zero.

As a consequence, the part of the utility function that captures the impact of the quality belief (see Equation 4) can be written as:

$$\text{(A.19)}\quad -\exp\left[-r_i\left(\bar{q}_{it}^{R} - r_i\left(s_{it}^{R}\right)^2/2\right)\right] = -\exp\left[-r_i\cdot\left(\bar{q}_0 + \bar{k}\cdot\sum_{\tau=1}^{L^R}\max(0,T-t-\tau)/L^R\right)\right.$$
$$\left. + r_i^2\cdot\left((.01)^2 + \left(s^d\right)^2 + \left(\sum_{\tau=1}^{L^R}\max(0,T-t-\tau)/L^R\right)^2\cdot\left(s^k\right)^2\right)/2\right].$$

The to-be-estimated parameters in this expression are $\bar{q}_0$, $\bar{k}$, s^k, s^d, $\overline{r^*}$ and σ^{r^*} (the latter two being the parameters of the heterogeneity distribution of the risk aversion parameter r_i). Notice that the functional form of Expression A.19 and the fixed value of s_0 (= .01) identify the scale and location of the heterogeneous risk aversion parameter r_i. That is, it is not possible to nullify changes in $\overline{r^*}$ or σ^{r^*} by compensating changes in the other parameters. Moreover, because $\sum_{\tau=1}^{L^R}\max(0,T-t-\tau)/L^R$ changes as a function of time, we can also disentangle $\bar{q}_0$, $\bar{k}$, s^d, and s^k.

Identification of the remaining coefficients in the utility equations. Separate identification of the informative and persuasive effects of advertising and direct marketing stems from the fact that, while the persuasive impact on customer utility at a given point in time is fixed, the informative effect is not: a customer that received more signals in the past, has become more certain about the service's ultimate quality, and learns less from new incoming advertising and direct marketing signals. Hence, the portion of the marketing communication impact that dies down as the consumer learns tends to be the informative effect, and the portion that persists relates to the persuasive effect (Mehta, Chen, & Narasimhan, 2008b).

The constant term and the coefficients of the (free and paid) trial dummies can be separated from the exponential CARA expression because, for a given consumer, they contribute to utility in the same way at any point in time, whereas the role of the CARA expression depends on the number of quality signals received and the passage of time. The coefficients of the two dummy variables that capture preceding trial (free or paid) do not interfere with the CARA expression either, because their effect persists even when a consumer's usage during the trial (and thus learning through usage) is limited.

The time trend in the utility functions can be separated from any trending patterns in the learning part, because it has a fixed utility effect at any point in time, irrespective of the type and number of signals received by a customer. Finally, the coefficients of the household characteristics (age, income, household size, and relationship length), the uncertainty reduction variable, the price-discount variable, and the competitive advertising variable can be identified easily on the basis of observed variation in these variables.

Importantly, next to analyzing the intuition underlying the parameter identification, we also constructed an artificial dataset on the basis of the original independent variables and found that our code was able to recover the true relationships with a reasonable degree of accuracy.

Web Appendix 4: Log Likelihood Function of Main Model

$$\text{(A.20)}\quad LL = \sum_{i=1}^{10,000} \ln\Bigg[\int\left(\prod_{t=1}^{T_i}\left(P_{it}^{F}\right)^{\upsilon_{it}^{F}}\left(P_{it}^{R}\right)^{\upsilon_{it}^{R}}\left(P_{it}^{N}\right)^{\upsilon_{it}^{N}}\right)$$

$$\cdot f\left(\beta_i^0\middle|\alpha^0,\alpha^Z;\sigma^{\beta^0}\right)f\left(\beta_i\middle|\overline{\beta};\sigma^{\beta}\right)f\left(r_i^*\middle|\overline{r^*};\sigma^{r^*}\right)f\left(\theta_i\middle|\overline{\theta};\sigma^{\theta}\right)$$

$$\cdot f\left(\vartheta_i^A,\vartheta_i^D,\vartheta_i^U,\vartheta_i^W\middle|\overline{d^*},\kappa,\overline{k};\sigma^A,\sigma^D,\sigma^U,\sigma^W\right)d\left(\beta_i^0,\beta_i,r_i^*,\theta_i,\vartheta_i^A,\vartheta_i^D,\vartheta_i^U,\vartheta_i^W\right)\Bigg]$$

where T_i is the number of observations for consumer i; υ_{it}^{F}, υ_{it}^{R}, and υ_{it}^{N} indicate whether the consumer chose the free trial, the regular offer, or the no-purchase option in month t; β_i refers to the vector of consumer-specific response coefficients, except for β_i^0 (see Equations 1, 2, and 15); ϑ_i^A is a vector containing consumer i's (unconditional) corrected advertising signal sums $\vartheta_{i1}^A,\dots,\vartheta_{iT_i}^A$; similarly, $\vartheta_i^D = \vartheta_{i1}^D,\dots,\vartheta_{iT_i}^D$, $\vartheta_i^U = \vartheta_{i1}^U,\dots,\vartheta_{iT_i}^U$, and $\vartheta_i^W = \vartheta_{i1}^W,\dots,\vartheta_{iT_i}^W$; $\overline{\beta}$ is the vector of population means for the consumer-specific response coefficients β_i; $\overline{r^*}$ and $\overline{\theta}$ are the population means of the consumer-specific parameters r_i^* and θ_i; $\sigma^{\beta^0}, \sigma^{\beta}, \sigma^{r^*}$, and σ^{θ} are the population-level standard deviations of $\beta_i^0, \beta_i, r_i^*$, and θ_i, respectively; and f(.) refers to a normal density function. The other symbols are defined in the text.

Web Appendix 5: Robustness of Usage, WOM, Advertising, and Direct Marketing Effects

A possible concern is that the quality signals are not exogenous or pick up latent phenomena which may bias the results. First, usage may depend on perceived quality such that consumers use the service more intensively when quality beliefs improve. This could result in an overestimation of the usage signals' effects. Because quality beliefs improve over time, usage endogeneity would likely surface in the form of increasing usage rates. Yet, if anything, there appears to be a minor *decrease* in usage intensity: the correlation between usage and the trend variable equals −.038. Usage endogeneity therefore does not seem to be a major concern in our context.

Second, the number of WOM signals (i.e., the number of subscribers surrounding a given consumer) may be a proxy for certain omitted consumer characteristics that correlate with service subscription, if people living in the same area are akin. However, we explicitly control for the effects of several important socio-demographics; moreover, our measurement of WOM is not constrained to a limited geographic band around each consumer but also captures the influence of more remote, and possibly very different, consumers. Another concern is that the estimated effect of WOM could be biased by network externalities when service quality improves as a function of the number of adopters, for example, due to increased viewer participation in live TV shows. In the studied setting, however, this sort of user interaction is virtually nonexistent because the provider does not *produce* TV content.

Third, advertising, which in principle is region-specific, could capture unobserved region-specific demand shifts. However, the variation in advertising spending across regions turns out to be negligible such that this risk in practice does not exist.

Fourth, direct marketing could be cross-sectionally correlated with the error term of the adoption equation due to targeting. However, because the firm is not likely to target its communication based on (full) information about the individuals' propensity to adopt, but, rather, on household characteristics like socio-demographics or relationship length that we control for, we do not expect this to be a problem (see Narayanan & Manchanda, 2009 for a similar argument). Still, to check for this possible source of endogeneity, we followed Mundlak (1978) and Risselada, Verhoef, and Franses (2014) and included the consumer-specific average direct marketing level as a covariate. The pattern of results (in particular the informative and persuasive effects of direct marketing) remained largely unchanged.

References

Ching, A. T., Erdem, T., & Keane, M. P. (2011). *Learning Models: An Assessment of Progress, Challenges and New Developments*. Working paper. Rotman School of Management, University of Toronto. Toronto.Mehta, N., Chen, X. J., & Narasimhan, O. (2008). Informing, Transforming, and Persuading: Disentangling the Multiple Effects of Advertising on Brand Choice Decisions. *Marketing Science, 27*(3), 334-355.

Ching, A. T., Erdem, T., & Keane, M. P. (2013). Learning Models: An Assessment of Progress, Challenges, and New Developments. *Marketing Science, 32*(6), 913-938.

Coscelli, A, & Shum, M. (2004). An empirical model of learning and patient spillovers in new drug entry. *Journal of Econometrics*, *122*(2), 213–246.

Mehta, N., Chen, X. J., & Narasimhan, O. (2008a). Informing, Transforming, and Persuading: Disentangling the Multiple Effects of Advertising on Brand Choice Decisions. *Marketing Science, 27*(3), 334-355.

Mehta, N., Chen, X. J., & Narasimhan, O. (2008b). Technical Appendix to Accompany "Informing, Transforming, and Persuading: Disentangling the Multiple Effects of Advertising on Brand Choice Decisions." *Marketing Science, 27*(3).

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69-85.

Narayanan, S., & Manchanda, P. (2009). Heterogeneous Learning and the Targeting of Marketing Communication for New Products. *Marketing Science, 28*(3), 424-441.

Risselada, H., Verhoef, P. C., & Franses, P. H. (2014). Dynamic Effects of Social Influence and Direct Marketing on the Adoption of High-Technology Products. *Journal of Marketing, 78*(2), 52-68.

Weiss, Neil A. (2005), *A Course in Probability 1st ed.*, Pearson.