A. Proof of Proposition 1

With regard to the mixture parameters, τ, we use Proposition 1 of Teicher (1963) that states that a finite mixture of Gaussians is identifiable besides index relabeling, which in our case is a mixture of log-Gaussians. Thus, the only identifiability we have left to show is with respect to the parameter θ. Consider any mixture τ for the download distribution and $\theta^1 \neq \theta^2$. Assume by contradiction that for every possible instance, it holds that $ll(\tau, \theta^1) = ll(\tau, \theta^2)$. Let us denote $I(\theta)$ the last term to the right of Equation (1) (we omit its dependence on τ as it does not play any relevant role for our purpose). Note it must hold that $I(\theta^1) = I(\theta^2)$; otherwise, both likelihoods, $ll(\tau, \theta^1)$ and $ll(\tau, \theta^2)$, would not equate. Consider the infinite set of download samples parametrized by $y > 0$, where $\{d_i^{(y)}\}_{i=1}^N$ such that $d_1^{(y)} = y$, and $d_i^{(y)} = d_i^{(z)}$ for all $i > 1$ and for all $z \neq y$. Equating the log-likelihoods for θ^1 and θ^2 for any fixed values of $y, z > 0$ $(y \neq z)$ leads to $\sum_{i=1}^N \ln\left(1 - \Phi\left(l^0(d_i^{(y)})|\gamma(\theta^1, d_i^{(y)})\right)\right) = \sum_{i=1}^N \ln\left(1 - \Phi\left(l^0(d_i^{(y)})|\gamma(\theta^2, d_i^{(y)})\right)\right)$, and similarly for z. We proceed by subtracting the LHS of the likelihoods (with θ^1) of y and z, and equating with the subtraction of the RHS of the likelihood (with θ^2) which leads to $\ln(1 - \Phi(l^0(y)|\gamma(\theta^1, y))) - \ln(1 - \Phi(l^0(z)|\gamma(\theta^1, z))) = \ln(1 - \Phi(l^0(y)|\gamma(\theta^2, y))) - \ln(1 - \Phi(l^0(z)|\gamma(\theta^2, z)))$. Since this holds for any $y, z > 0$, then we have that $\frac{d}{dx}(\ln(1 - \Phi(l^0(x)|\gamma(\theta^1, x)))) = \frac{d}{dx}(\ln(1 - \Phi(l^0(x)|\gamma(\theta^2, x))))$. Note that $l^0(x)$ is differentiable almost everywhere (as there are some finite points in which there is no derivative), so it must be the case that $\ln(1 - \Phi(l^0(x)|\gamma(\theta^1, x))) = \ln(1 - \Phi(l^0(x)|\gamma(\theta^2, x))) + C$ for some constant $C \in \mathbb{R}$. Take by Assumption 1 some $d_0 > 0$ such that $\gamma(\theta^1, d_0) > \gamma(\theta^2, d_0)$ (wlog. due to relabeling θ^1 and θ^2). Since $\ln(1 - \Phi(l^0(x)|\gamma))$ is a strictly increasing function of γ (this will be shown in the end of the proof), then $\ln(1 - \Phi(l^0(d_0)|\gamma(\theta^1, d_0))) > \ln(1 - \Phi(l^0(d_0)|\gamma(\theta^2, d_0)))$ which implies $C > 0$, then $\ln(1 - \Phi(l^0(x)|\gamma(\theta^1, x))) = \ln(1 - \Phi(l^0(x)|\gamma(\theta^2, x))) + C$ for all $x > 0$. This last leads to $ll(\tau, \theta^1) > ll(\tau, \theta^2)$, which results in a contradiction. Let us show the strictly increasing behavior of $\ln(1 - \Phi(l^0|\gamma))$ as a function of γ. Note that $\Phi(l^0|\gamma) = \Phi(l^0 \times (\kappa\gamma)^{-1} - \kappa^{-1}|0, 1)$, that is, a standard normal. Since functions, $\ln(\cdot)$, $1 - \Phi(\cdot|0, 1)$, and $l^0 \times (\kappa\gamma)^{-1} - \kappa^{-1}$, are so that only the first function is strictly increasing and the last two are strictly decreasing (where γ is the argument of the last function), then $\ln(1 - \Phi(l^0|\gamma))$ is strictly increasing in γ, which concludes the proof.

B. Constant expressions in bounding the log-likelihood

The expressions for terms C_1 and C_2 are

$$
\begin{aligned}
C_1 &= \sum_{i \in \mathcal{I}_o} \ln\left(1 - \Phi\left(l^0(d_i)|\gamma(\theta, d_i)\right)\right) - \sum_{i \in \mathcal{I}_o} \ln(d_i) \\
C_2 &= -\sum_{i \in \mathcal{I}_o} \sum_{j=1}^{J} p_i(j) \ln(p_i(j)) + m \int_0^\infty \sum_{j=1}^{J} p(j, x) \ln\left(\frac{\Phi\left(l^0(x)|\gamma(\theta, x)\right)}{p(j, x) x}\right) dx.
\end{aligned}
$$

C. Derivation of Expressions (3) and (4) of the EM algorithm

We provide step by step the details on the computation of the expressions for $p_i(j)$ and $p(j,x)$. As mentioned in the main body of the paper, consider a fixed parameter $\tau^{(k)} = \{(\mu_j^{(k)}, \sigma_j^{(k)}, \eta_j^{(k)})\}_{j=1}^J$, where k represents a particular iteration. Then

$$
\begin{aligned}
p_i(j) &= \mathbb{P}(Z_i = j | D_i = d_i, W_i \geq l^0(D_i), \tau^{(k)}) \\
&= \frac{\mathbb{P}(D_i = d_i, W_i \geq l^0(D_i) | Z_i = j, \tau^{(k)}) \mathbb{P}(Z_i = j | \tau^{(k)})}{\sum_{h=1}^J \mathbb{P}(D_i = d_i, W_i \geq l^0(D_i) | Z_i = h, \tau^{(k)}) \mathbb{P}(Z_i = h | \tau^{(k)})} \\
&= \frac{\mathbb{P}(W_i \geq l^0(D_i) | D_i = d_i, Z_i = j, \tau^{(k)}) \mathbb{P}(D_i = d_i | Z_i = j, \tau^{(k)}) \mathbb{P}(Z_i = j | \tau^{(k)})}{\sum_{h=1}^J \mathbb{P}(W_i \geq l^0(D_i) | D_i = d_i, Z_i = h, \tau^{(k)}) \mathbb{P}(D_i = d_i | Z_i = h, \tau^{(k)}) \mathbb{P}(Z_i = h | \tau^{(k)})} \\
&= \frac{\mathbb{P}(W_i \geq l^0(D_i) | D_i = d_i, \tau^{(k)}) \mathbb{P}(D_i = d_i | Z_i = j, \tau^{(k)}) \mathbb{P}(Z_i = j | \tau^{(k)})}{\sum_{h=1}^J \mathbb{P}(W_i \geq l^0(D_i) | D_i = d_i, \tau^{(k)}) \mathbb{P}(D_i = d_i | Z_i = h, \tau^{(k)}) \mathbb{P}(Z_i = h | \tau^{(k)})} \\
&= \frac{\phi(\ln(d_i) | \mu_j^{(k)}, \sigma_j^{(k)}) \eta_j^{(k)} / \ln(d_i)}{\sum_{h=1}^J \phi(\ln(d_i) | \mu_h^{(k)}, \sigma_h^{(k)}) \eta_h^{(k)} / \ln(d_i)} \\
&= \frac{\phi(\ln(d_i) | \mu_j^{(k)}, \sigma_j^{(k)}) \eta_j^{(k)}}{\sum_{h=1}^J \phi(\ln(d_i) | \mu_h^{(k)}, \sigma_h^{(k)}) \eta_h^{(k)}}, \\
p(j,x) &= \mathbb{P}(Z = j, D = x | W < l^0(D), \tau^{(k)}) \\
&= \frac{\mathbb{P}(W < l^0(D) | D = x, Z = j, \tau^{(k)}) \mathbb{P}(D = x | Z = j, \tau^{(k)}) \mathbb{P}(Z = j | \tau^{(k)})}{\int_0^\infty \sum_{h=1}^J \mathbb{P}(W < l^0(D) | D = y, Z = h, \tau^{(k)}) \mathbb{P}(D = y | Z = h, \tau^{(k)}) \mathbb{P}(Z = h | \tau^{(k)}) dy} \\
&= \frac{\mathbb{P}(W < l^0(D) | D = x, \tau^{(k)}) \mathbb{P}(D = x | Z = j, \tau^{(k)}) \mathbb{P}(Z = j | \tau^{(k)})}{\int_0^\infty \sum_{h=1}^J \mathbb{P}(W < l^0(D) | D = y, \tau^{(k)}) \mathbb{P}(D = y | Z = h, \tau^{(k)}) \mathbb{P}(Z = h | \tau^{(k)}) dy} \\
&= \frac{\Phi(l^0(x) | \gamma(\theta, x)) \phi(\ln(x) | \mu_j^{(k)}, \sigma_j^{(k)}) \eta_j^{(k)} x^{-1}}{\int_0^\infty \sum_{h=1}^J \Phi(l^0(y) | \gamma(\theta, y)) \phi(\ln(y) | \mu_h^{(k)}, \sigma_h^{(k)}) \eta_h^{(k)} y^{-1} dy}
\end{aligned}
$$

the first equality is the definition of $p_i(j)$. The second equality is Bayes' rule and Total Law of Probability in the denominator. In the fourth equality, we drop the conditioning events $(Z_i = j)$ from $\mathbb{P}(D_i = d_i, W_i \geq l^0(D_i) | Z_i = j, \tau^{(k)})$ since it does not provide any additional information as we already have the precise download value in the conditioning terms, similarly in the analogous expression in the denominator. In the fifth inequality, the term $\mathbb{P}(W_i \geq l^0(D_i) | D_i = d_i, \tau^{(k)})$ is simplified since it appears in the numerator and denominator as well.

D. Integrals computed in Algorithm 3.1

Algorithm 3.1 requires the computation of several integrals. These are $\int_0^\infty p(j,x)dx$, $\int_0^\infty p(j,x)\ln(x)dx$, $\int_0^\infty p(j,x)(\ln(x) - \mu)^2 dx$ for all $j \in \{1, \ldots, J\}$, the integrals in Equation (2), and those in Equation (12). The arguments of the integrals are complex terms in the sense of computing the integrals in closed form. Note that $p(j,x)$ is the CDF of a Normal distribution, where its argument, mean, and standard deviation are nonlinear functions of the integration variable x. Standardizing to a 0–1 Normal distribution does not make things better, since the argument of the CDF (which would be $(l_0(x) - \mu(\theta, x))/(k \cdot \mu(\theta, x))$) would still be a nontrivial function of the integration variable x. In practice, we compute these integrals in Julia programming language using the function "quadgk" of the package "QuadGK". The latter function uses the adaptive Gauss-Kronrod quadrature method. This is used in an adaptive way to compute estimates of the integral over intervals, so that the interval with the most error is subdivided into two halves. The Gauss-Kronrod method is of particular interest since it reuses the points used for the Gaussian quadrature, thereby reducing the number of computations. In spite of the difficulties of finding closed form expressions, the numerical integration is computed reasonably quickly, because of the fact that it is a one-dimensional integral (although there are CDFs in the integral arguments which, by definition, are also integrals). The CDFs, are computed using the built-in functions of the package "StatsFun" (also in Julia), such as the function "normcdf" of the "StatsFun" package.

E. Relation between Prices and Downloads

We consider x a vector of downloads defined as $x = \{d \in \mathbb{R} | \exists s \in \{0, 1, \ldots, |\mathcal{S}|\}, \exists l \in \mathbb{Z} \cap [1, 10], d = l(b_s - b_{s-1})/10\}$ with $b_0 = 0$, namely, x is a set of download values so that for each pair of plans with consecutive index, we take ten download values even-spaced. Denote $y = l(x)$, where $l(\cdot)$ is defined as $l(x, \mathbf{f^0}, \mathbf{v^0}) = \min_{s \in \mathcal{S}} f_s + v_s \times \max\{x - b_s, 0\}$. The following linear regressions are performed between: $y \sim x$, $\ln(y) \sim x$, $y \sim \ln(x)$, and $\ln(y) \sim \ln(x)$. All regressions have constant term. The R^2 obtained from each regression, in the same order as they are presented, are: 0.713, 0.922, 0.376, and 0.966. The log-log relationship between customers' payments and download is the one that has the highest coefficient of determination, thus we decided to choose a log-log relationship between the willingness to pay (actually a statistic of the willingness to pay since this is a random variable such as the mean), and the download amount.

F. Log-likelihood for Gaussian WTP Different Mean Reservation Price Functions

See Tables 1 and 2.

G. Log-likelihood for Uniform and Exponential WTP Distributions for different Mean Reservation Price Functions

For the uniform distribution, we consider that the WTP follows a distribution in which the mean follows a function $\gamma(\theta, d)$ of a vector of parameters θ and download value d, and, similarly as the Gaussian case, we consider different coefficients of variations $\kappa \in \{1/4, 1/3, 1/2\}$. As before, for each functional form γ, we run Algorithm 3.1 to estimate distribution parameters τ and θ on a training set, and evaluate the log-likelihood attained in the testing set. This is done using 5-fold cross-validation. Results are shown in Tables 3, 4, and 5.

$\ln(\gamma(\theta,d))$	$\nu=1/3$	$\nu=2/3$	$\nu=1$	$\nu=4/3$	$\nu=5/3$
$\theta_1+\theta_2\ln(d)$	−39914.55	−42164.97	−43768.99	−45025.97	−46067.02
$\theta_1+\theta_2\ln(d+1)$	−39905.48	−42157.15	−43762.88	−45020.69	−46058.43
$\theta_1+\theta_2\ln(d+\theta_3)$	−39905.78	−42157.21	−43762.05	−45018.78	−46055.48
$\theta_1+\theta_2\ln(d)+\theta_3\ln(d)^2$	−39906.22	−42161.52	−43767.01	−45024.12	−46064.11
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^2$	−39904.19	−42154.93	−43759.26	−45014.74	−46049.64
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$	−39903.44	−42151.07	−43753.16	−45008.30	−46043.34
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^{0.5}$	−39904.49	−42155.65	−43761.23	−45018.99	−46056.84
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^{0.5}$	−39904.74	−42153.53	−43756.29	−45012.09	−46048.33

Table 1 Average log-likelihood among the 5-fold cross-validation set for $\kappa=1/4$ for different parametric functions of the mean reservation price after running Algorithm 3.1.

$\ln(\gamma(\theta,d))$	$\nu=1/3$	$\nu=2/3$	$\nu=1$	$\nu=4/3$	$\nu=5/3$
$\theta_1+\theta_2\ln(d)$	−39904.88	−42149.30	−43753.81	−45010.68	−46077.65
$\theta_1+\theta_2\ln(d+1)$	−39903.24	−42143.54	−43740.24	−44991.74	−46024.15
$\theta_1+\theta_2\ln(d+\theta_3)$	−39900.70	−42141.79	−43739.60	−44991.37	−46024.23
$\theta_1+\theta_2\ln(d)+\theta_3\ln(d)^2$	−39899.41	−42141.08	−43740.12	−44992.65	−46027.10
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^2$	−39899.97	−42138.30	−43734.69	−44984.71	−46015.22
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$	−39898.59	−42138.32	−43734.19	−44984.41	−46015.75
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^{0.5}$	−39901.14	−42140.91	−43737.58	−44989.28	−46021.62
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^{0.5}$	−39899.82	−42139.75	−43736.70	−44988.50	−46020.69

Table 2 Average log-likelihood among the 5-fold cross-validation set for $\kappa=1/2$ for different parametric functions of the mean reservation price after running Algorithm 3.1.

$\ln(\gamma(\theta,d))$	$\nu=1/3$	$\nu=2/3$	$\nu=1$	$\nu=4/3$	$\nu=5/3$
$\theta_1+\theta_2\ln(d)$	−39914.85	−42468.31	−44827.25	−45319.49	−46468.50
$\theta_1+\theta_2\ln(d+1)$	−39904.25	−42145.56	−43752.68	−45011.85	−46056.83
$\theta_1+\theta_2\ln(d+\theta_3)$	−39904.38	−42144.83	−43741.25	−44998.26	−46033.11
$\theta_1+\theta_2\ln(d)+\theta_3\ln(d)^2$	−39910.59	−42160.24	−43790.58	−45049.64	−46106.43
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^2$	−39903.01	−42149.68	−43756.46	−45019.53	−46062.70
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$	−39901.85	−42150.96	−43757.40	−45019.38	−46065.74
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^{0.5}$	−39904.66	−42157.45	−43773.71	−45047.52	−46108.38
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^{0.5}$	−39903.47	−42148.90	−43761.88	−45020.52	−46072.19

Table 3 Average log-likelihood among the 5-fold cross-validation set with Uniform WTP for $\kappa=1/4$ for different parametric functions of the mean reservation price after running Algorithm 3.1.

H. Generation of Missing Data

For an observed customer i, her willingness to pay is generated according to the fitted distribution for WTP W but conditioned on the event that this customer is an observed one, i.e., she purchases, therefore $W_i \geq l^0(d_i)$. The density of this random variable can be computed as:

$$f_W(w|W_i \geq l^0(d_i)) = \frac{\mathbb{P}(W_i \geq l^0(d_i)|W_i = w) f_W(w)}{\int_{-\infty}^{\infty} \mathbb{P}(W_i \geq l^0(d_i)|W_i = s) f_W(s) ds} = \frac{\mathbb{1}_{\{w \geq l^0(d_i)\}} \phi(w|\gamma(\theta, d_i))}{1 - \Phi(l^0(d_i)|\gamma(\theta, d_i))}.$$

The first equality is due to Bayes' rule and the Total Probability Law.

For unobserved customers, their download is sampled according to the fitted download distribution but conditioned on the event that this customer does not purchases. The density of this

$\ln(\gamma(\theta,d))$	$\nu=1/3$	$\nu=2/3$	$\nu=1$	$\nu=4/3$	$\nu=5/3$
$\theta_1+\theta_2\ln(d)$	-39907.70	-42155.99	-44136.23	-45396.29	-46534.42
$\theta_1+\theta_2\ln(d+1)$	-39903.08	-42139.73	-43738.73	-44995.01	-46036.19
$\theta_1+\theta_2\ln(d+\theta_3)$	-39903.25	-42140.98	-43735.65	-44989.98	-46022.90
$\theta_1+\theta_2\ln(d)+\theta_3\ln(d)^2$	-39907.11	-42141.39	-43745.06	-45005.88	-46048.80
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^2$	-39902.19	-42141.02	-43739.69	-44993.86	-46032.81
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$	-39902.56	-42140.15	-43739.45	-44995.39	-46036.55
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^{0.5}$	-39902.30	-42139.91	-43739.17	-44994.85	-46036.56
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^{0.5}$	-39903.81	-42142.84	-43739.02	-44991.41	-46025.83

Table 4 Average log-likelihood among the 5-fold cross-validation set with Uniform WTP for $\kappa=1/3$ for different parametric functions of the mean reservation price after running Algorithm 3.1.

$\ln(\gamma(\theta,d))$	$\nu=1/3$	$\nu=2/3$	$\nu=1$	$\nu=4/3$	$\nu=5/3$
$\theta_1+\theta_2\ln(d)$	-39903.69	-42146.68	-43857.30	-45375.20	-46356.66
$\theta_1+\theta_2\ln(d+1)$	-39903.16	-42139.91	-43732.67	-44982.87	-46016.83
$\theta_1+\theta_2\ln(d+\theta_3)$	-39903.21	-42138.25	-43732.55	-44982.05	-46013.25
$\theta_1+\theta_2\ln(d)+\theta_3\ln(d)^2$	-39901.31	-42137.35	-43735.72	-44984.95	-46021.09
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^2$	-39901.37	-42138.65	-43732.77	-44983.92	-46018.85
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^2$	-39902.18	-42139.14	-43732.87	-44982.90	-46015.50
$\theta_1+\theta_2\ln(d+1)+\theta_3\ln(d+1)^{0.5}$	-39902.07	-42139.51	-43732.66	-44982.81	-46016.80
$\theta_1+\theta_2\ln(d+\theta_4)+\theta_3\ln(d+\theta_4)^{0.5}$	-39901.33	-42138.70	-43733.90	-44984.99	-46019.82

Table 5 Average log-likelihood among the 5-fold cross-validation set with Uniform WTP for $\kappa=1/2$ for different parametric functions of the mean reservation price after running Algorithm 3.1.

probability distribution function can be computed as follows:

$$\begin{aligned} f_D(x|W<l^0(D)) &= \frac{\mathbb{P}(W<l^0(D)|D=x)f_D(x)}{\int_0^\infty \mathbb{P}(W<l^0(D)|D=y)f_D(y)dy} \\ &= \frac{\Phi\left(l^0(x)|\gamma(\theta,x)\right)\sum_{j=1}^J \phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}}{\int_0^\infty \Phi\left(l^0(y)|\gamma(\theta,y)\right)\sum_{j=1}^J \phi(\ln(y)|\mu_j,\sigma_j)\eta_j y^{-1}dy}. \end{aligned}$$

The first equality is due to Bayes' rule and the Total Probability Law.

Finally, the way we generate the WTP for unobserved customers is analogous to the case of observed ones, except that we condition in the event that the customers does not purchase. The probability density function can be computed as:

$$f_W(w|W_i<l^0(d_i)) = \frac{\mathbb{P}(W_i<l^0(d_i)|W_i=w)f_W(w)}{\int_{-\infty}^\infty \mathbb{P}(W_i<l^0(d_i)|W_i=s)f_W(s)ds} = \frac{\mathbb{1}_{\{w<l^0(d_i)\}}\phi(w|\gamma(\theta,d_i))}{\Phi(l^0(d_i)|\gamma(\theta,d_i))}.$$

As a side note, it is worth noticing that for the case of observed customers, we sample their download values from the empirical distribution, i.e., from the empirical values we observe. Alternatively, we could also sample these values by using $D|W\geq l^0(D)$, which would also be technically correct.

I. Proof of Proposition 4

Let $(\mathbf{f}^*,\mathbf{v}^*)$ be the optimal solution of (11). Define $(\hat{\mathbf{f}}^*,\hat{\mathbf{v}}^*)$ such that $\hat{f}_t=\lceil f_t^*/\varepsilon_f\rceil\varepsilon_f$ and $\hat{v}_t=\lceil v_t^*/\varepsilon_v\rceil\varepsilon_v$. Note that this solution $(\hat{\mathbf{f}}^*,\hat{\mathbf{v}}^*)$ is feasible on the UBDP (of course, it might not be the optimal one of UBDP). Then

$$
\begin{aligned}
z^* &= \sum_{t=1}^{S}\sum_{i\in\mathcal{N}_t} l(D_i,\mathbf{f},\mathbf{v})\mathbb{1}_{\{W_i\geq l(D_i,\mathbf{f},\mathbf{v})\}}\\
&= f_1^*\sum_{i\in\mathcal{N}_1}\mathbb{1}_{\{W_i\geq f_1^*\}} + \sum_{t=2}^{S}\sum_{i\in\mathcal{N}_t}\min\{f_{t-1}^*+v_{t-1}^*e_{i,t-1},f_t^*\}\mathbb{1}_{\{\min\{W_i\geq f_{t-1}^*+v_{t-1}^*e_{i,t-1},f_t^*\}\}}\\
&\leq \hat{f}_1^*\sum_{i\in\mathcal{N}_1}\mathbb{1}_{\{W_i\geq(\hat{f}_t^*-\varepsilon_f)^+\}} + \sum_{t=2}^{S}\sum_{i\in\mathcal{N}_t}\min\{\hat{f}_{t-1}^*+\hat{v}_{t-1}^*e_{i,t-1},\hat{f}_t^*\}\mathbb{1}_{\{W_i\geq\min\{(\hat{f}_{t-1}^*-\varepsilon_f)^+ +(\hat{v}_{t-1}^*-\varepsilon_v)^+e_{i,t-1},(\hat{f}_t^*-\varepsilon_f)^+\}\}}\\
&\leq V_1'^*(s^1) = z^{UBDP},
\end{aligned}
$$

where the first inequality is because $(\hat{f}_t^*-\varepsilon_f)^+\leq f_t^*\leq\hat{f}^*$ and $(\hat{v}_t^*-\varepsilon_v)^+\leq v_t^*\leq\hat{v}^*$. Indeed, $(\hat{f}_t^*-\varepsilon_f)^+\leq f_t^*$ is equivalent to $(\lceil x\rceil-1)^+\leq x$ for $x\geq 0$. Since this last condition is clearly true for $x\in[0,1]$, for $x>1$ this is equivalent to $\lceil x\rceil-1\leq x$, which always holds. With respect to $f_t^*\leq\hat{f}^*$, this follows immediately by the definition of $\hat{f}^*$. Similarly is the case for the inequalities of v_t^*. The second inequality is because the revenue attained on the UBDP by the path solution followed by $(\hat{\mathbf{f}}^*,\hat{\mathbf{v}}^*)$ as $a^t=(\hat{f}_t^*,\hat{v}_{t-1}^*)$ for $t>1$ and $a^1=\hat{f}_1^*$ is, at most, equal to the optimal revenue z^{UBDP}.

J. Proof of Lemma 1

For any N, it holds that $\mathbb{E}[z_N]\geq z^*$ (see Proposition 5.6 of Shapiro et al. (2009)), so it follows naturally that $\mathbb{E}[z_N^{UBDP}]\geq z^*$ since $z_N^{UBDP}\geq z_N$ for any random variable realization. Then, a one-sided confidence interval can be built on z_N^{UBDP} to give an upper bound with $1-\alpha$ confidence on $\mathbb{E}[z_N^{UBDP}]$, which by transitivity also applies to z^*. This interval can be simply given by the sample average plus a multiple (depending on the confidence level $1-\alpha$) of an unbiased estimate of the standard deviation. More precisely, given that there are M samples of the optimization problem in which the upper bound $z_N^{UBDP,m}$ is computed for each sample m, the upper bound is given by $\overline{z}_{N,M}^{UBDP}+t_{\alpha,M-1}\hat{\sigma}_{N,M}^2$ where $\overline{z}_{N,M}^{UBDP}:=\frac{1}{M}\sum_{m=1}^{M}z_N^{UBDP,m}$, $\hat{\sigma}_{N,M}:=\frac{1}{M}\left(\frac{1}{M-1}\sum_{m=1}^{M}\left(z_N^{UBDP,m}-\overline{z}_{N,M}^{UBDP}\right)^2\right)$, and $t_{alpha,M-1}$ is the CDF at $1-\alpha$ of a t-distribution with $M-1$ degrees of freedom.

K. Expectation over the Assumed Parameters

An alternative approach would be to consider a distribution over the distributions parameters. For example, we can assume that with probability 1/3 customers' WTP coefficient of variation (κ') is either 1/4, 1/3, or 1/2, and similarly with the ratio of unobserved customers ν'. In other words, the expectation of Optimization Problems (10) and (11) can also be taken over a distribution of the assumed problem parameters. We consider a discrete uniform distribution for κ' (ν') on $\{1/4,1/3,1/2\}$ ($\{1/3,2/3,1,4/3,5/3\}$). Figure 1 shows the percentage revenue difference of the optimal DPP prices versus the current ones for different cases, when considering the expectation on the SAA with respect to either customers' WTP coefficient of variation (κ'), the ratio of unobserved customers (ν'), or both. Although there are cases with revenue losses, these are not as big as

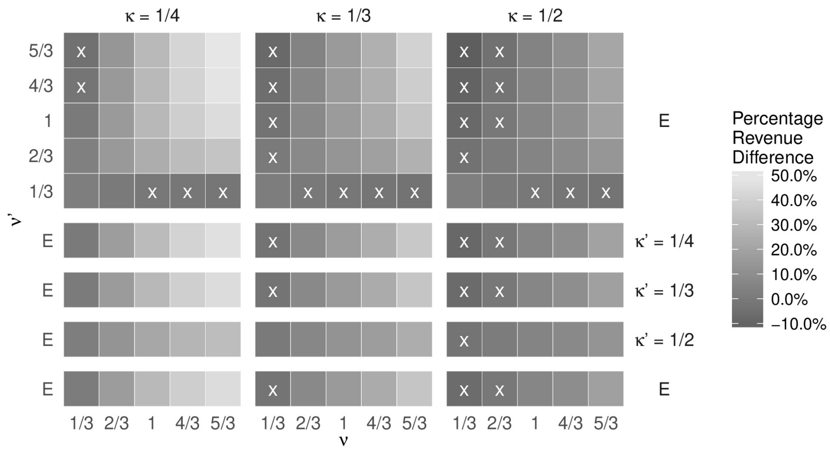

Figure 1 **Revenue difference is computed as $\frac{z^{DPP}_{(\kappa,\nu),(\kappa',\nu')} - z^{0}_{(\kappa,\nu)}}{z^{0}_{(\kappa,\nu)}}$. Assumed parameters (κ', ν') are in rows. "E" indicate the cases where the assumed parameter (κ' or ν') follow a discrete uniform distribution. Real parameters (κ, ν) are in columns. Cases where the revenue difference is negative are marked with "$\times$".**

in the worst cases depicted in Section 5.3. On the one hand, if the expectation is taken just on customers' WTP coefficient of variation, under- and overestimation of unobserved customers is the main driver whether or not revenues increase. Indeed, in the case when assuming $\nu' = 1/3$, and thus few unobserved customers, revenues tend to decrease in almost every case (see the fifth row of Figure 1) due to overpricing. On the other hand, if the expectation is taken only on the ratio of unobserved customers, revenues are increased in almost all cases, with few exceptions when there are few unobserved customers in reality (lower values of ν), and therefore under-pricing occurs. Interestingly, the particular case when considering $\kappa' = 1/2$ (and the expectation for ν') appears to be a robust strategy where there are no revenue losses in any scenario (i.e., a column).

From the analysis of Sections 5.3 and K, taking expectation over the parameters' distributions helps to reduce the risk in revenue losses. Although there can be some adverse cases in reality, there are several in which revenues can be significantly improved. Finally, it is worth noting that the previous analysis highlights the importance of getting unknown information. For instance, if we know through a market study that consumers' WTP is low, then taking expectation over the considered scenarios of the number of unobserved customers seems to be a good way to go.

L. Price-dependent consumption

The way to model consumption dependent on price is going to be captured by a utility function that depends on the download as well as on the payment. Similar to Lambrecht et al. (2007), we consider that a customer has a reference download amount that represents the ideal consumption at a null price. Let us denote as $r_i > 0$ this amount for each customer i, which is unobserved, yet it is known by the user. Then the utility of customer i is expressed as:

$$u_i(d) = w_i - \frac{1}{2c}(\ln(d) - \ln(r_i))^2 - l(d, \mathbf{f}, \mathbf{v}). \tag{1}$$

The first term, w_i represents the consumer individual valuations which they know but are not observable to us. Let W be the random variable of the latter amount. The parameters of the distribution of W depend on the reference download, and a vector of parameters θ. The second term represents a disutility for deviating from the reference download amount. $c > 0$ captures the sensitivity of the utility with respect to deviations of the reference download. Note that the previous model, which was introduced without consumption being dependent on price, can be represented if (i) $c \to 0$ and $r_i = d_i$, i.e., the only feasible consumption of a consumer is the one that matches her reference value. The third term corresponds to the payment of the cheapest plan. If we consider a single plan, i.e., $l(d) = f + v\max\{d - b, 0\}$, the maximizer of problem (1) results in (i) r_i if $r_i \le b$, (ii) $r_i - vc$ if $b < r_i - vc$, and (iii) b if $r_i - vc \le b < r_i$. In reality, we do not observe download values equal to the bundle amount b, therefore we introduce a noise uncertainty into the consumption. More precisely, we consider inducing uncertainty in the download, d_i, such that the actual values distribute according to d_iY_i, where Y_i is a log-Normal distribution with standard deviation of $a > 0$ and mean $-a^2/2$ so that $\mathbb{E}[Y_i] = 1$. Wlog, for ease of handling the mathematical expressions, we add the term $-a^2/(2c)$ to the consumer utility. Then, the expected consumer utility of purchasing plan s can be expressed as:

$$\begin{aligned}
&e_{is}(d)\\
&= w_i - \mathbb{E}\left[\frac{1}{2c}\ln\left(\frac{dY_i}{r_i}\right)^2\right] - \frac{a^2}{2c} - f_s - v_s\mathbb{E}\left[\max\{dY_i - b_s, 0\}\right]\\
&= w_i - \left(\frac{1}{2c}\left(\ln\left(\frac{d}{r_i}\right) - \frac{a^2}{2}\right)^2 + f_s + dv_s\mathbb{E}\left[\max\left\{Y_i - \frac{b_s}{d}, 0\right\}\right]\right)\\
&= w_i - \left(\frac{1}{2c}\left(\ln\left(\frac{d}{r_i}\right) - \frac{a^2}{2}\right)^2 + f_s - v_sb_s + dv_s\mathbb{E}\left[\max\left\{Y_i, \frac{b_s}{d}\right\}\right]\right)\\
&= w_i - \left(\frac{1}{2c}\left(\ln\left(\frac{d}{r_i}\right) - \frac{a^2}{2}\right)^2 + f_s - v_sb_s + dv_s\frac{b_s}{d}\Phi\left(\frac{\ln\left(\frac{b_s}{d}\right) + \frac{a^2}{2}}{a}\right) + dv_s\Phi\left(\frac{\frac{a^2}{2} - \ln(\frac{b_s}{d})}{a}\right)\right)\\
&= w_i - \underbrace{\left(\frac{1}{2c}\left(\ln\left(\frac{d}{r_i}\right) - \frac{a^2}{2}\right)^2 + f_s - v_sb_s\left(1 - \Phi\left(\frac{\ln\left(\frac{b_s}{d}\right) + \frac{a^2}{2}}{a}\right)\right) + dv_s\Phi\left(\frac{\frac{a^2}{2} - \ln(\frac{b_s}{d})}{a}\right)\right)}_{\lambda(d,s,r_i,a,c)}. \quad (2)
\end{aligned}$$

In the second equality we have used the fact that if $X \sim \text{Lognormal}(\mu, \sigma^2)$, then for any $f > 0$, $fX \sim \text{Lognormal}(\mu + \ln(f), \sigma^2)$, and thus $\log(fX) \sim \text{Normal}(\mu + \ln(f), \sigma^2)$. Then $\mathbb{E}[\ln(fX)^2] = \sigma^2 + \mathbb{E}[X]^2 = \sigma^2 + \mu^2$. Replacing $f = d/r_i$ and $Y_i = X$, leads to $\mathbb{E}[\ln(dY_i/r_i)^2] = a^2 + (\ln(\frac{d}{r_i}) - \frac{a^2}{2})$. In the fourth equality we have used the fact that for a random variable $X \sim \text{Lognormal}(\mu, \sigma^2)$ and a scalar $x > 0$, $\mathbb{E}[X^2] = e^{2\mu + 2\sigma^2}$ and $\mathbb{E}[\max\{X, x\}] = x\Phi(\frac{\ln(x) - \mu}{\sigma}) + e^{\mu + \frac{\sigma^2}{2}}\Phi\left(\frac{\mu + \sigma^2 - \ln(x)}{\sigma}\right)$. It is worth mentioning that for the plan with unlimited bundle data, the expression $\lambda(d,s,r_i,a,c)$ consists only of the first term above the brace shown in Equation (2). Unless stated otherwise, the expressions shown are for plans that have finite bundle data, i.e., $b_s<\infty$. The first order condition of Equation (2) can also be computed as:

$$\frac{\partial e_{is}}{\partial d}=\frac{-1}{cd}\left(\ln\left(\frac{d}{r_i}\right)-\frac{a^2}{2}\right)+\frac{v_sb_s}{da}\phi\left(\frac{\frac{a^2}{2}+\ln\left(\frac{b_s}{d}\right)}{a}\right)-v_s\Phi\left(\frac{\frac{a^2}{2}-\ln(\frac{b_s}{d})}{a}\right)-\frac{v_s}{a}\phi\left(\frac{\frac{a^2}{2}-\ln(\frac{b_s}{d})}{a}\right) \quad (3)$$

Let us call d^*_{is} the maximizer of Equation (2). This clearly exists since (i) the function is continuous in $(0,+\infty)$, (ii) λ goes to $+\infty$ when $d\to 0^+$, and (iii) λ goes to $+\infty$ when $d\to+\infty$. (i) is direct. For (ii), note that the first of λ goes to infinity when $d\to 0^+$, the second term is constant, the third term is bounded below by $-v_sb_s$, and the fourth term goes to 0. For (iii), note that the first term of λ goes to ∞, the second term is constant, the third term is bounded above by $-v_sb_s$, and the last term is non-negative.

We assume a customer prefers the plan that reports the highest expected utility, namely, $s^*_i=\arg\max_{s\in\mathcal{S}} e_{is}(d^*_{is})$. Also, let us define the corresponding utility of purchasing this plan as $v_i=e_{is^*_i}(d^*_{is^*_i})$. For ease of notation we, denote the optimal download value as $d^*_i=d^*_{is^*_i}$. Then, the consumer is observed if and only if $t_i>0$. Note that the maximizer download of Equation (2) among all possible plan options can be written as $d^*(x)$ as a function of the reference download x. Similarly, we define

$$\lambda^*(x)=\min_{d>0,s\in\mathcal{S}}\lambda(d,s,x,a,c). \quad (4)$$

For ease of notation, we omit the dependence for a given pair of parameters (a,c).

Let R_i be the random variable to denote the reference download which will be distributed according to a mixture of Gaussians with parameters $\tau=\{(\mu_j,\sigma_j,\eta_j)\}_{j=1}^J$. The random variable that represents the actual download of an observed customer is denoted as D_i. Given the value of R_i, we can obtain the download value d^*_i from solving Problem (2), and thus $D_i=d^*_iY_i$. Then, we denote the PDF of $D|R$ as $f_{D|R}$. As before, we assume that the set of observed (unobserved) customers is denoted by $\mathcal{I}_o$ ($\mathcal{I}_u$) and $|\mathcal{I}_o|=n$ ($|\mathcal{I}_u|=m$) Then, the log-likelihood can be written as

$$\begin{aligned}
&ll(\tau,\theta,a,c)\\
=&\sum_{i\in\mathcal{I}_o}\ln\left(\mathbb{E}_{R_i}\left[\mathbb{P}(t_i>0,D_i=d_i|R_i)\right]\right)+\sum_{i\in\mathcal{I}_u}\ln\left(\mathbb{E}_{R_i}\left[\mathbb{P}(t_i\le 0|R_i)\right]\right)\\
=&\sum_{i\in\mathcal{I}_o}\ln\left(\mathbb{E}_{R_i}\left[\mathbb{P}(t_i>0|D_i=d_i,R_i)f_{D|R}(d_i)\right]\right)+m\ln\left(\mathbb{E}_R\left[\mathbb{P}(t\le 0|R)\right]\right)\\
=&\sum_{i\in\mathcal{I}_o}\ln\left(\mathbb{E}_{R_i}\left[\mathbb{P}(t_i>0|R_i)f_{D|R}(d_i)\right]\right)+m\ln\left(\mathbb{E}_R\left[\mathbb{P}(t\le 0|R)\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\in\mathcal{I}_o} \ln\left(\int_0^\infty \mathbb{P}(t_i>0|R_i=x) f_{D|R=x}(d_i) \sum_{j=1}^{J} \frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{x} dx\right) \\
&\quad + m\ln\left(\int_0^\infty \mathbb{P}(t\le 0|R=x) \sum_{j=1}^{J} \frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{x} dx\right) \\
&= \sum_{i\in\mathcal{I}_o} \ln\left(\int_0^\infty \left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right) \frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)} \sum_{j=1}^{J} \frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{x} dx\right) \\
&\quad + m\ln\left(\int_0^\infty \Phi\left(\lambda^*(x)|\gamma(\theta,x)\right) \sum_{j=1}^{J} \frac{\phi(\ln(x)|\mu_j,\sigma_j)\eta_j}{x} dx\right) \qquad (5)
\end{aligned}
$$

Optimization on τ: A key idea of the EM algorithm is to lower bound the log-likelihood function. Denote Z_i as the random variable that represents the mixture where the reference download of observed customer i belongs, and Z as the analogous random variable for an unobserved customer. Unlike observed customers, unobserved customers are indistinguishable. Thus, we define a single representative random variable for them. Also consider $p_i(j,x)$ as the probability that an observed customer i belongs to mixture j and has a reference downloads of x, and that $p(j,x)$ is the probability density function that an unobserved customer belongs to mixture j, and has a reference download of x. Then, the log-likelihood, as a function of τ, can be expressed as follows:

$$
\begin{aligned}
&ll(\tau) \\
&= \sum_{i\in\mathcal{I}_o} \ln\left(\int_0^\infty \left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right) \frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)} \sum_{j=1}^{J} \frac{\phi(\ln(x)|\mu_j,\sigma_j)}{x} dx\right) \\
&\quad + m\ln\left(\int_0^\infty \Phi\left(\lambda^*(x)|\gamma(\theta,x)\right) \sum_{j=1}^{J} \frac{\phi(\ln(x)|\mu_j,\sigma_j)}{x} dx\right) \\
&= \sum_{i\in\mathcal{I}_o} \ln\left(\int_0^\infty \sum_{j=1}^{J} p_i(j,x) \left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right) \frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)} \frac{\phi(\ln(x)|\mu_j,\sigma_j)}{xp_i(j,x)} dx\right) \\
&\quad + m\ln\left(\int_0^\infty \sum_{j=1}^{J} p(j,x)\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right) \frac{\phi(\ln(x)|\mu_j,\sigma_j)}{xp(j,x)} dx\right) \\
&\ge \sum_{i\in\mathcal{I}_o} \int_0^\infty \sum_{j=1}^{J} p_i(j,x) \ln\left(\left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right) \frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)} \frac{\phi(\ln(x)|\mu_j,\sigma_j)}{xp_i(j,x)} dx\right) \\
&\quad + m\int_0^\infty \sum_{j=1}^{J} p(j,x) \ln\left(\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right) \frac{\phi(\ln(x)|\mu_j,\sigma_j)}{xp(j,x)} dx\right) \\
&= \sum_{i\in\mathcal{I}_o} \int_0^\infty \sum_{j=1}^{J} p_i(j,x) \ln\left(\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\right) dx + C_1 \\
&\quad + m\int_0^\infty \sum_{j=1}^{J} p(j,x) \ln\left(\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\right) dx + C_2 \qquad (6)
\end{aligned}
$$

Where

$$C_1 = \sum_{i\in\mathcal{I}_o}\int_0^\infty \sum_{j=1}^J p_i(j,x)\ln\left((1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right))\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\frac{1}{xp_i(j,x)}dx\right)$$
$$C_2 = m\int_0^\infty \sum_{j=1}^J p(j,x)\ln\left(\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\frac{1}{xp(j,x)}dx\right)$$

The inequality follows from Jensen's inequality.

- **E-Step:** We compute the expressions for $p_i(j,x)$ and $p(j,x)$ so that they match the posterior probability of the unobserved data given the observed and assumed information for a fixed parameter, $\tau^{(k)} = \{(\mu_j^{(k)},\sigma_j^{(k)},\eta_j^{(k)})\}_{j=1}^J$, where k represents a particular iteration. In order to reduce the notation, we omit the parameter dependence on $\tau^{(k)}$.

$$\begin{aligned}
&= p_i(j,x) = \mathbb{P}(Z_i=j,R_i=x|v_i>0,D_i=d_i)\\
&= \frac{\mathbb{P}(t_i>0,D_i=d_i|R_i=x,Z_i=j)\mathbb{P}(R_i=x,Z_i=j)}{\int_0^\infty\sum_{h=1}^J \mathbb{P}(t_i>0,D_i=d_i|R_i=y,Z_i=h)\mathbb{P}(R_i=y,Z_i=h)dy}\\
&= \frac{\mathbb{P}(t_i>0|D_i=d_i,R_i=x,Z_i=j)\mathbb{P}(D_i=d_i|R_i=x,Z_i=j)\mathbb{P}(R_i=x|Z_i=j)\mathbb{P}(Z_i=j)}{\int_0^\infty\sum_{h=1}^J \mathbb{P}(t_i>0|D_i=d_i,R_i=y,Z_i=h)\mathbb{P}(D_i=d_i|R_i=y,Z_i=j)\mathbb{P}(R_i=y|Z_i=h)\mathbb{P}(Z_i=h)dy}\\
&= \frac{\mathbb{P}(t_i>0|R_i=x)\mathbb{P}(D_i=d_i|R_i=x)\mathbb{P}(R_i=x|Z_i=j)\mathbb{P}(Z_i=j)}{\int_0^\infty\sum_{h=1}^J \mathbb{P}(t_i>0|R_i=y)\mathbb{P}(D_i=d_i|R_i=y)\mathbb{P}(R_i=y|Z_i=h)\mathbb{P}(Z_i=h)dy}\\
&= \frac{(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right))\,\phi(\ln(d_i/d^*(x))|-\frac{a^2}{2},a^2)d^*(x)d_i^{-1}\phi(\ln(x)|\mu_j^{(k)},\sigma_j^{(k)})x^{-1}\eta_j^{(k)}}{\int_0^\infty\sum_{h=1}^J (1-\Phi\left(\lambda^*(y)|\gamma(\theta,y)\right))\,\phi(\ln(d_i/d^*(y))|-\frac{a^2}{2},a^2)d^*(y)d_i^{-1}\phi(\ln(y)|\mu_h^{(k)},\sigma_h^{(k)})y^{-1}\eta_h^{(k)}dy}
\end{aligned} \tag{7}$$

$$\begin{aligned}
p(j,x) &= \mathbb{P}(Z=j,R=x|t\le 0) = \frac{\mathbb{P}(t\le 0|R=x,Z=j)\mathbb{P}(R=x,Z=j)}{\int_0^\infty\sum_{h=1}^J \mathbb{P}(t\le 0|R=y,Z=h)\mathbb{P}(R=y,Z=h)dy}\\
&= \frac{\mathbb{P}(t\le 0|R=x,Z=j)\mathbb{P}(R=x|Z=j)\mathbb{P}(Z=j)}{\int_0^\infty\sum_{h=1}^J \mathbb{P}(t\le 0|R=y,Z=h)\mathbb{P}(R=y|Z=h)\mathbb{P}(Z=h)dy}\\
&= \frac{\mathbb{P}(t\le 0|R=x)\mathbb{P}(R=x|Z=j)\mathbb{P}(Z=j)}{\int_0^\infty\sum_{h=1}^J \mathbb{P}(t\le 0|R=y)\mathbb{P}(R=y|Z=h)\mathbb{P}(Z=h)dy}\\
&= \frac{\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\phi(\ln(x)|\mu_j^{(k)},\sigma_j^{(k)})x^{-1}\eta_j^{(k)}}{\int_0^\infty\sum_{h=1}^J \Phi\left(\lambda^*(y)|\gamma(\theta,y)\right)\phi(\ln(y)|\mu_h^{(k)},\sigma_h^{(k)})y^{-1}\eta_h^{(k)}dy}
\end{aligned} \tag{8}$$

- **M-Step:** Setting the FOC to zero with respect to μ_j, σ_j, and η_j on the Lagrangian of Equation (6) (namely Equation (6) plus the term $\lambda\left(\sum_{j=1}^J \eta_j - 1\right)$ due to the constraint $\sum_{j=1}^J \eta_j = 1$), leads to:

$$\mu_j = \frac{\sum_{i\in\mathcal{I}_o}\int_0^\infty p_i(j,x)\ln(x)dx + m\int_0^\infty p(j,x)\ln(x)\,dx}{\sum_{i\in\mathcal{I}_o}\int_0^\infty p_i(j,x)dx + m\int_0^\infty p(j,x)dx} \tag{9}$$

$$\sigma_j^2 = \frac{\sum_{i\in\mathcal{I}_o}\int_0^\infty p_i(j,x)(\ln(x)-\mu_j)^2dx + m\int_0^\infty p(j,x)(\ln(x)-\mu_j)^2\,dx}{\sum_{i\in\mathcal{I}_o}\int_0^\infty p_i(j,x)dx + m\int_0^\infty p(j,x)dx} \tag{10}$$

$$\eta_j = \frac{\sum_{i\in\mathcal{I}_o}\int_0^\infty p_i(j,x)dx + m\int_0^\infty p(j,x)dx}{n+m}. \tag{11}$$

Optimization on θ: We consider the gradient ascent method with respect to θ while keeping τ fixed. Taking the FOC of Equation (5) with respect to θ_i gives rise to:

$$\begin{aligned}
&\frac{\partial}{\partial\theta_k} ll(\tau,\theta,a,c)\\
&= \sum_{i\in\mathcal{I}_o}\frac{\int_0^\infty \phi(\lambda^*(x)|\gamma(\theta,x))\frac{\lambda^*(x)}{\gamma(\theta,x)}\frac{\partial\gamma(\theta,x)}{\partial\theta_k}\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right)\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}\\
&-\frac{m\int_0^\infty \phi(\lambda^*(x)|\gamma(\theta,x))\frac{\lambda^*(x)}{\gamma(\theta,x)}\frac{\partial\gamma(\theta,x)}{\partial\theta_k}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}.
\end{aligned} \tag{12}$$

In the case of the derivative of the log-likelihood with respect to parameter a, we have

$$\begin{aligned}
&\frac{\partial}{\partial a} ll(\tau,\theta,a,c)\\
&= \sum_{i\in\mathcal{I}_o}\frac{\int_0^\infty -\phi(\lambda^*(x)|\gamma(\theta,x))\frac{\partial\lambda^*(x)}{\partial a}\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right)\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}\\
&+\sum_{i\in\mathcal{I}_o}\frac{\int_0^\infty\left(1-\Phi(\lambda^*(x)|\gamma(\theta,x))\right)\frac{\partial}{\partial a}\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right)\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}\\
&+\frac{m\int_0^\infty \phi(\lambda^*(x)|\gamma(\theta,x))\frac{\partial\lambda^*(x)}{\partial a}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}.
\end{aligned}$$

For ease of notation we will omit the integration variable x on the following. Let $w_i := \frac{\ln(d_i)-\ln(d^*)+\frac{a^2}{2}}{a}$. Also, we use $\phi()$ to denote the PDF of a standard Gaussian. We need to compute $\frac{\partial}{\partial a}\frac{d^*\phi(w_i)}{ad_i}$ and $\frac{\partial\lambda^*}{\partial a}$. Then $\frac{\partial}{\partial a}\left(\frac{d^*}{d_ia}\phi(w_i)\right) = \frac{\partial d^*}{\partial a}\frac{\phi(w_i)}{ad_i} - \frac{d^*\phi(w_i)}{d_ia^2} - \frac{w_id^*\phi(w_i)}{d_ia}\frac{\partial w_i}{\partial a}$. Also, we have that $\frac{\partial w_i}{\partial a} = -\frac{1}{ad^*}\frac{\partial d^*}{\partial a} - \frac{1}{a^2}\ln\left(\frac{d_i}{d^*}\right)+\frac{1}{2}$. In order to obtain $\frac{\partial d^*}{\partial a}$, let us define $y := \frac{1}{a}\ln\left(\frac{b_{s^*}}{d^*}\right)+\frac{a}{2}$ and $z := -\frac{1}{a}\ln\left(\frac{b_{s^*}}{d^*}\right)+\frac{a}{2}$. Note that $\frac{\partial y}{\partial a} = -\frac{1}{ad^*}\frac{\partial d^*}{\partial a}+\frac{z}{a}$ and $\frac{\partial z}{\partial a} = \frac{1}{ad^*}\frac{\partial d^*}{\partial a}+\frac{y}{a}$. Taking the derivative of Equation (3) with respect to a equating to zero, replacing $\frac{\partial y}{\partial a}$ and $\frac{\partial z}{\partial a}$, leads to

$$\frac{\partial d^*}{\partial a} = \frac{\ln\left(\frac{x}{d^*}\right)+\frac{3}{2}a^2 - \frac{cv_{s^*}b_{s^*}y\phi(y)z}{a} - cd^*v_{s^*}\Phi(z) - ycd^*v_{s^*}\phi(z) + \frac{cv_{s^*}d^*y\phi(z)z}{a}}{\frac{a}{d^*} - \frac{cv_{s^*}b_{s^*}y\phi(y)}{ad^*} + acv_{s^*}\Phi(z) + 2cv_{s^*}\phi(z) - \frac{cv_{s^*}z\phi(z)}{a}} \tag{13}$$

For the case of the unlimited plan, i.e., $b_{s^*}=\infty$, the only term of Equation (3) that remains is the first one, namely we would have $-(\ln(\frac{d^*}{x})-\frac{a^2}{2})/(cd)=0$, and equivalently $d^*=xe^{\frac{a^2}{2}}$. Taking the derivative of the latter expression at both sides with respect to a leads to $\frac{\partial d^*}{\partial a} = axe^{\frac{a^2}{2}}$.

To obtain $\frac{\partial\lambda^*(x)}{\partial a}$, we can use the Envelope Theorem and get $\frac{\partial\lambda^*(x)}{\partial a} = -\frac{a}{c}\left(\ln\left(\frac{d^*}{x}\right)-\frac{a^2}{2}\right) + v_{s^*}b_{s^*}\phi(y)\frac{z}{a} + d^*v_{s^*}\phi(z)\frac{y}{a}$. For the unlimited plan, i.e., $b_{s^*}=\infty$, $\lambda(d^*,s^*,x,a,c) = (\ln(\frac{d^*}{x})-\frac{a^2}{2})^2/(2c)$, the partial derivative with respect to a equals $-\frac{a}{c}\left(\ln\left(\frac{d^*}{x}\right)-\frac{a^2}{2}\right)$.

Now we look for the log-likelihood derivative with respect to parameter c.

$$\begin{aligned}
&\frac{\partial}{\partial c} ll(\tau,\theta,a,c) \\
=&\sum_{i\in\mathcal{I}_o}\frac{\int_0^\infty -\phi(\lambda^*(x)|\gamma(\theta,x))\frac{\partial\lambda^*(x)}{\partial c}\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right)\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx} \\
&+\sum_{i\in\mathcal{I}_o}\frac{\int_0^\infty\left(1-\Phi(\lambda^*(x)|\gamma(\theta,x))\right)\frac{\partial}{\partial c}\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\left(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\right)\frac{\phi(\ln(d_i/d^*(x))|-a^2/2,a^2)}{d_i/d^*(x)}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx} \\
&+\frac{m\int_0^\infty\phi(\lambda^*(x)|\gamma(\theta,x))\frac{\partial\lambda^*(x)}{\partial c}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j\, x^{-1}dx}{\int_0^\infty\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)\eta_j x^{-1}dx}.
\end{aligned}$$

Where $\frac{\partial}{\partial c}\frac{\phi(\ln(d_i/d^*)|-a^2/2,a^2)}{d_i/d^*} = \frac{\partial}{\partial c}\left(\frac{d^*}{d_i a}\phi(w_i)\right) = \frac{\partial d^*}{\partial c}\frac{\phi(w_i)}{ad_i} - \frac{w_i d^*\phi(w_i)}{d_i a}\frac{\partial w_i}{\partial c}$. We also have that $\frac{\partial w_i}{\partial c} = -\frac{1}{ad^*}\frac{\partial d^*}{\partial c}$. To obtain $\frac{\partial d^*}{\partial c}$, recall the definitions of the auxiliary variables y and z, we have that $\frac{\partial y}{\partial c} = -\frac{1}{ad^*}\frac{\partial d^*}{\partial c}$ and $\frac{1}{ad^*}\frac{\partial d^*}{\partial a}$. Taking the derivatives at both sides of Equation (3) with respect to c, and replacing the derivatives of y and z, we obtain

$$\frac{\partial d^*}{\partial c} = \frac{v_{s^*}b_{s^*}\phi(y) - ad^*v_{s^*}\Phi(z) - d^*v_{s^*}\phi(z)}{\frac{a}{d^*} - \frac{cv_{s^*}b_{s^*}y\phi(y)}{ad^*} + acv_{s^*}\Phi(z) + 2cv_{s^*}\phi(z) - \frac{cv_{s^*}z\phi(z)}{a}} \tag{14}$$

For the unlimited plan, i.e., $b_{s^*} = \infty$, the only term of Equation (3) that remains is the first one, namely we would have $-(\ln(\frac{d^*}{x}) - \frac{a^2}{2})/(cd) = 0$, and equivalently $d^* = xe^{\frac{a^2}{2}}$. Taking the derivatives of the latter expression at both sides with respect to c leads to $\frac{\partial d^*}{\partial c} = 0$.

To obtain $\frac{\partial\lambda^*(x)}{\partial c}$, we can use the Envelope Theorem and get $\frac{\partial\lambda^*(x)}{\partial c} = -\frac{1}{2c^2}\left(\ln\left(\frac{d^*}{x}\right) - \frac{a^2}{2}\right)^2$ where in this case, we are deriving only partially with respect to c (because of the Envelope Theorem).

The full log-likelihood maximization algorithm is summarized below:

Algorithm L.1: Estimate parameters of the probability distributions of customers' download and willingness to pay

Input : ν, κ, $\gamma(\theta,d):\mathbb{R}^u\times\mathbb{R}_+\to\mathbb{R}_+$, ϵ, ϵ_τ, ϵ_θ
Output: $\tau=\{(\mu_j,\sigma_j,\eta_j)\}_{j=1}^J$, θ, a, c

Initialize $\theta^{(0)}$, $a^{(0)}$, $c^{(0)}$, and $\tau^{(0)}$. $h=1$, $ll^{(0)}=-\infty$, $ll^{(h)}=ll(\tau^{(0)},\theta^{(0)},a^{(0)},c^{(0)})$;
while $ll^{(h)}-ll^{(h-1)}>\epsilon$ **do**
    /* Optimization on τ */
    $ll_\tau^{(0)}=-\infty$, $ll_\tau^{(1)}=ll^{(h)}$, $(\theta,a,c)=(\theta^{(k-1)},a^{(k-1)},c^{(k-1)})$, $k=1$;
    while $ll_\tau^{(k)}-ll_\tau^{(k-1)}>\epsilon_\tau$ **do**
        /* E-Step */
        $k=k+1$;
        Compute $p_i(j,x)$ from Equation (7), for all $i\in\{1,\dots,n\}$;
        Use $p(j,x)$ from Equation (8) to compute $\int_0^\infty p(j,x)dx$, $\int_0^\infty p(j,x)\ln(x)dx$, and $\int_0^\infty p(j,x)(\ln(x)-\mu_j)^2dx$, for all $j\in\{1,\dots,J\}$;
        /* M-Step */
        Compute $\tau^{(k)}$ ($\mu_j^{(k)}$, $\sigma_j^{(k)}$, and $\eta_j^{(k)}$) from Equations (9), (10), and (11);
        $ll_\tau^{(k)}=ll(\tau^{(k)},\theta,a,c)$;
    /* Optimization on θ, a, and c */
    $ll_\theta^{(0)}=-\infty$, $ll_\theta^{(1)}=ll_\tau^{(k)}$, $\tau=\tau^{(k)}$, $k=1$;
    while $ll_\theta^{(k)}-ll_\theta^{(k-1)}>\epsilon_\theta$ **do**
        $k=k+1$;
        $(\theta^{(k)},a^{(k)},c^{(k)})=\theta^{(k-1)}+\beta\cdot\nabla_{(\theta,a,c)}ll(\tau,\theta^{(k-1)},a^{(k-1)},c^{(k-1)})$;
        $ll_\theta^{(k)}=ll(\tau,\theta^{(k)},a^{(k)},c^{(k)})$;
    $h=h+1$;
    $ll^{(h)}=ll_\theta^{(k)}$;

M. Generation of missing data

The procedure for generating the missing data in the case of price-dependent consumption is analogous to that presented in Section 3.5, except that in this case instead of sampling downloads, we have to sample customers' reference downloads from which we obtain the optimal download value for each of them. Then, we apply the noise component by multiplying the lognormal random variable to this download amount, obtaining the actual download consumption of the customer.

We proceed by bootrstrapping on a set of $1,000$ scenarios by running Algorithm L.1, obtaining estimates of the parameters τ, θ, a, and c for each scenario. As before, let N be the total number of customers for which we want to sample data. N represents the size of the sampled instance; it can take any arbitrary positive integer value. Each customer is considered to be an observed one with probability $1/(1+\nu)$ and an unobserved one with probability $\nu/(1+\nu)$. Then the generation of missing data is as follows:

For the observed customers

I Sample their downloads according to their empirical distribution obtained from the bootstrap.

Then, sample their reference downloads considering the information we obtain of their observed download amount. The PDF of this distribution to sample is:

$$f_{R_i}(r|D_i=d_i,t_i>0)=\frac{\mathbb{P}(t_i>0|R_i=r)\mathbb{P}(D_i=d_i|R_i=r)f_{R_i}(r)}{\int_0^\infty \mathbb{P}(t_i>0|R_i=x)\mathbb{P}(D_i=d_i|R_i=x)f_{R_i}(x)\,dx}$$
$$=\frac{(1-\Phi\left(\lambda^*(r)|\gamma(\theta,r)\right))\,\phi(\ln(d_i/d^*(r))|-\frac{a^2}{2},a^2)d^*(r)d_i^{-1}\sum_{j=1}^J\phi(\ln(r)|\mu_j,\sigma_j)r^{-1}\eta_j}{\int_0^\infty(1-\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right))\,\phi(\ln(d_i/d^*(x))|-\frac{a^2}{2},a^2)d^*(x)d_i^{-1}\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)x^{-1}\eta_j\,dx}$$

II Sample their willingness to pay from the distribution with PDF

$$f_{W_i}(w|D_i=d_i,t_i>0)=\int_0^\infty \mathbb{P}(W_i=w|R_i=x,t_i>0)\mathbb{P}(R_i=x|D_i=d_i,t_i>0)\,dx$$
$$=\int_0^\infty \frac{\mathbb{P}(t_i>0|W_i=w,R_i=x)f_{W_i|R_i=x}(w)}{\mathbb{P}(t_i>0|R_i=x)}\mathbb{P}(R_i=x|D_i=d_i,t_i>0)\,dx$$
$$=\int_0^\infty \frac{\mathbb{1}_{\{w>\lambda^*(x)\}}f_{W_i|R_i=x}(w)}{\mathbb{P}(t_i>0|R_i=x)}\frac{\mathbb{P}(t_i>0|D_i=d_i,R_i=x)\mathbb{P}(D_i=d_i|R_i=x)f_{R_i}(x)\,dx}{\int_0^\infty\mathbb{P}(t_i>0|D_i=d_i,R_i=y)\mathbb{P}(D_i=d_i|R_i=y)f_{R_i}(y)\,dy}$$
$$=\frac{\int_0^\infty \mathbb{1}_{\{w>\lambda^*(x)\}}f_{W_i|R_i=x}(w)\mathbb{P}(D_i=d_i|R_i=x)f_{R_i}(x)\,dx}{\int_0^\infty\mathbb{P}(t_i>0|R_i=y)\mathbb{P}(D_i=d_i|R_i=y)f_{R_i}(y)\,dy}$$
$$=\frac{\int_0^\infty \mathbb{1}_{\{w>\lambda^*(x)\}}\phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\phi(\ln(d_i/d^*(x))|-\frac{a^2}{2},a^2)d^*(x)d_i^{-1}f_{R_i}(x)\,dx}{\int_0^\infty(1-\Phi\left(\lambda^*(y)|\gamma(\theta,y)\right))\,\phi(\ln(d_i/d^*(y))|-\frac{a^2}{2},a^2)d^*(y)d_i^{-1}f_{R_i}(y)\,dy}$$

where $f_{R_i}(x)=\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)x^{-1}\eta_j$.

For the unobserved customers

III Sample their downloads from the distribution with PDF

$$f_R(r|t\le 0)=\frac{\mathbb{P}(t\le 0|R=l)f_R(r)}{\int_0^\infty\mathbb{P}(t\le 0|R=x)f_R(x)\,dx}=\frac{\Phi\left(\lambda^*(r)|\gamma(\theta,r)\right)\sum_{j=1}^J\phi(\ln(r)|\mu_j,\sigma_j)r^{-1}\eta_j}{\int_0^\infty\Phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)x^{-1}\eta_j\,dx}$$

IV Sample their willingness to pay from the distribution with PDF

$$f_W(w|t\le 0)=\int_0^\infty \mathbb{P}(W=w,R=x|t\le 0)\,dx=\int_0^\infty\frac{\mathbb{P}(t\le 0|W=w,R=x)\mathbb{P}(W=w,R=x)}{\mathbb{P}(t\le 0)}\,dx$$
$$=\frac{\int_0^\infty\mathbb{1}_{\{w\le\lambda^*(x)\}}\phi\left(\lambda^*(x)|\gamma(\theta,x)\right)\sum_{j=1}^J\phi(\ln(x)|\mu_j,\sigma_j)x^{-1}\eta_j\,dx}{\int_0^\infty\Phi\left(\lambda^*(y)|\gamma(\theta,y)\right)\sum_{j=1}^J\phi(\ln(y)|\mu_j,\sigma_j)y^{-1}\eta_j\,dy}$$

N. Sample Average Approximation

We describe how to perform SAA with the pricing problem faced by the seller. Consider the same definition for λ^* that was defined in Equation (4), but making the dependency on the fixed and variable prices explicit, namely, $\lambda^*(r,\mathbf{f},\mathbf{v})$ where $r>0$ is a reference download value. Then, for an observed customer i (under the current prices $(\mathbf{f^0},\mathbf{v^0})$), it holds that $w_i>\lambda^*(r_i,\mathbf{f^0},\mathbf{v^0})$. The fraction of observed and unobserved customers is $1/(1+\nu)$ and $\nu/(1+\nu)$ respectively. Then, the seller objective function can be written as

$$\sum_{i\in\mathcal{I}_o}\mathbb{E}_{W_i,R_i}\left[\mathbb{E}_{Y_i}\left[f_{s_i^*}+v_{s_i^*}\max\{Y_id_i^*-b_{s_i^*},0\}|W_i>\lambda^*(R_i,\mathbf{f^0},\mathbf{v^0}),D_i=d_i\right]\mathbb{1}_{\{W_i>\lambda^*(R_i,\mathbf{f},\mathbf{v})\}}\right]+$$

$$\sum_{i\in\mathcal{I}_u} \mathbb{E}_{W_i,R_i}\left[\mathbb{E}_Y\left[f_{s_i^*}+v_{s_i^*}\max\{Y_i d_i^* - b_{s_i^*},0\}|W_i \le \lambda^*(R_i,\mathbf{f^0},\mathbf{v^0})\right]\mathbb{1}_{\{W_i>\lambda^*(R_i,\mathbf{f},\mathbf{v})\}}\right]$$
$$= n(1+\nu)\left(\mathbb{E}_{W,R,D}\left[\frac{1}{1+\nu}\mathbb{E}_Y\left[f_{s^*}+v_{s^*}\max\{Yd^*-b_{s^*},0\}|W>\lambda^*(R_i,\mathbf{f^0},\mathbf{v^0}),D\right]\mathbb{1}_{\{W>\lambda^*(R,\mathbf{f},\mathbf{v})\}}\right]+\right.$$
$$\left.\mathbb{E}_{W,R}\left[\frac{\nu}{1+\nu}\mathbb{E}_Y\left[f_{s^*}+v_{s^*}\max\{Yd^*-b_{s^*},0\}|W\le\lambda^*(R,\mathbf{f^0},\mathbf{v^0})\right]\mathbb{1}_{\{W>\lambda^*(R,\mathbf{f},\mathbf{v})\}}\right]\right)$$

where D is the rv that represents the observed download from an observed customer, which is distributed according to the empirical distribution of downloads. Then, considering that a customer is observed with probability $1/(1+\nu)$ (i.e., $W>\lambda^*(R,\mathbf{f^0},\mathbf{v^0})$), and unobserved with probability $\nu/(1+\nu)$ (i.e., $W<\lambda^*(R,\mathbf{f^0},\mathbf{v^0})$), the pricing optimization problem can be expressed as

$$\max_{\mathbf{f},\mathbf{v}\in\Omega}\mathbb{E}_{W,R,D}\left[\mathbb{E}_Y\left[f_{s^*}+v_{s^*}\max\{Yd^*-b_{s^*},0\}\right]\mathbb{1}_{\{W>\lambda^*(R,\mathbf{f},\mathbf{v})\}}\right]. \tag{15}$$

Then, we solve sampled instances of (15) with the procedure described in Section M. Unlike when consumption is independent of the price, we cannot formulate this problem with dynamic programming. Instead, we use a heuristic solving approach to find a good solution. We describe how we solve sampled instances with a nearest-neighbor heuristic in the next section in the appendix.

O. Nearest-Neighbor

Consider the set $\Omega_{(\epsilon_f,\epsilon_v)}$ as the grid of fixed and variable prices in the set Ω, such that fixed and variable prices are a multiple of ϵ_f and ϵ_v, respectively. Given that we are at a fixed and variable price $(\mathbf{f}^\diamond,\mathbf{v}^\diamond)$ in $\Omega_{(\epsilon_f,\epsilon_v)}$, we define the neighboring solutions as the fixed and variable prices in which all components are equal to $(\mathbf{f}^\diamond,\mathbf{v}^\diamond)$ except for one price component. This price component that is different can be for either a fixed price, or a variable price. Namely, the neighbors of the prices $(\mathbf{f}^\diamond,\mathbf{v}^\diamond)$ are defined as $\mathcal{N}((\mathbf{f}^\diamond,\mathbf{v}^\diamond))=\{(\mathbf{f},\mathbf{v})\in\Omega_{(\epsilon_f,\epsilon_v)}|f_s=f_s^\diamond+\Delta_{fs}\cdot\epsilon_f, v_s=v_s^\diamond+\Delta_{vs},\Delta_{fs},\Delta_{vs}\in\{-1,1\},\sum_j(|\Delta_{fs}|+|\Delta_{vs}|)=1,(\mathbf{f},\mathbf{v})\in\Omega\}$. Then, at each iteration where we are at the current solution $(\mathbf{f}^\diamond,\mathbf{v}^\diamond)\in\Omega_{(\epsilon_f,\epsilon_v)}$, we evaluate the objective function in all the neighboring solutions of the set $\mathcal{N}((\mathbf{f}^\diamond,\mathbf{v}^\diamond))$. If the neighbor point with the highest revenue achieves a higher revenue than the current point, we move to the neighboring point and iterate again. If otherwise, i.e., there is no revenue improvement in any of the neighbors, we stop. Note that the procedure ends in a finite number of steps, since the set $\Omega_{(\epsilon_f,\epsilon_v)}$ is finite.

References

Lambrecht, Anja, Katja Seim, Bernd Skiera. 2007. Does uncertainty matter? consumer behavior under three-part tariffs. *Marketing Science* **26**(5) 698–710.

Shapiro, Alexander, Darinka Dentcheva, Andrzej Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.

Teicher, Henry. 1963. Identifiability of finite mixtures. *The Annals of Mathematical Statistics* 1265–1269.