WEB APPENDIX

W-A. FULL MODEL SPECIFICATION

Notation

i: user, *i* = 1,…,N; *t*: time, t = 1,…,T_i;

k: latent class, *k* = 1,…,K; *j*: advertising publisher, j = 1,...,J

y_{it}: the number of visits an user *i* makes to the site at time *t*

ϕ_{1it} : the probability of an user *i* inclass *k* of visiting the site once at time *t*

ϕ_{2itk} : the probability of an user *i* inclass *k* of visiting the site more than once at time *t*

AdStock_{it}: AdStock for an user *i* at time *t*, the mean-centered log of geometrically-smoothed publisher exposures and each publisher's differential effect.

x_{itj} : the geometrically-smoothed index of each publisher's exposures (ι_{ijt} with a carryover parameter$\delta \in [0,1]$.

z_{it} : the standardized number of days that have elapsed since the last exposure at time *t*

v_{it} : the mean-centered binary variable that indicates whether a user has visited the site within a week before time *t*

w_i : the standardized variables for a user *i*'s browsing behavior such as frequency of web usage, breadth of different websites, and the tendency to visit media, financial, and e-commerce sites.

C : a N-vector that indicates class allocation, with each element taking on a value of $\{1, \ldots, K\}$

Full model

$$(y_{it} | C_i = k, \phi_{1itk}, \phi_{2it} \ , \mu_{itk}, \tau_k) \ =$$

$$(1 - \phi_{1itk} - \phi_{2itk})^{I(y_{it}=0)} \phi_{1itk}{}^{I(y_{it}=1)} \ [\phi_{2itk} p_2(y_{it} | \mu_{itk}, \tau_k)]]^{I(y_{it}>1)}$$

$$\phi_{qitk} = \frac{\exp(V_{qk} + u_{iq})}{\sum_{r=0}^{2} \exp(V_{rk} + u_{ir})} \text{ for } q \in \{0,1,2\} \text{ and}$$

$$\log(\mu_{itk}) = V_{3k} + u_{i3} \text{ where}$$

$$V_{qk} = \alpha_{q1k} + \alpha_{q2k} A_{it} + \alpha_{q3k} A_{it}^2 + \alpha_{q4k} z_{it} + \alpha_{q5k} z_{it}^2 + \alpha_{q6k} v_{it} + \alpha_{q7k} v_{it} A_{it} + \alpha_{q8k} v_{it} z_{it}$$

$$\text{for} \quad q \in \{1, 2, 3\}$$

$$u_i | C_i \sim N_3(0, \Sigma_k)$$

$$\text{AdStock}_{it} \ = \ \log \left[\left(\sum_{j=1}^{J} \lambda_j \, x_{ijt} \right) + 1 \right] - \bar{x},$$

$$x_{ijt} = (1-\delta)\iota_{ijt} + \delta \cdot x_{ijt} \quad C_i \sim Cat(m_{i1}, \dots, m_{iK})$$

$$m_{ik} = \frac{\exp(w_i\gamma_k)}{\sum_{j=1}^{K} \exp(w_i\gamma_j)}$$

Likelihood

$$L(y|\{\alpha_k\},\{\Sigma_k\},\{u_i\},\{\lambda_j\},C,\{\gamma_k\}) \propto \prod_{i=1}^{N}\{\sum_{k=1}^{K} f(m_{ik}|\gamma_k,w_i)\,[\prod_{t=1}^{T_i} f\big(y_{it}\big|\alpha_k,\lambda_j,u_i,\tau_k\big)f(u_i|\Sigma_k)]\}$$

Priors

$$\alpha_{1k} \sim N(\mu_{\alpha 1} = 0, \Lambda_{\alpha 1} = 100\mathrm{I}_{2r})$$

$\alpha_{2k} \sim N(\mu_{\alpha 2} = 0, \Lambda_{\alpha 2} = 100\mathrm{I}_{r})$where *r* is the number of parameters

$$\tau_k \sim IG(n_0 = 0.1, \delta_0 = 0.1)$$

$$\pi(\Sigma_k) \sim IW(\nu_0 = 3, D_0 = I_3)$$

$\gamma \sim N(\mu_\gamma = 0, V_\gamma = 100I_w)$, where *w* is the number of parameters in browsing behavior

$$\lambda_j \sim \mathrm{Gamma}(\mathrm{shape} = \frac{1}{\theta}, \mathrm{scale} = \frac{1}{\theta})$$

$$\theta \sim Unif\left(0, \infty\right)$$

W-B: TREATMENT FOR MISSING ONLINE BEHAVIOR VARIABLES

Since the adverting agency manages two datasets (ads-focused impression dataset and publishers-based ad network dataset) separately, only when users who were exposed to target ads and visited the target advertiser's site (when it is part of the ad network), their unique identifiers are synchronized and allow us to observe their behaviors in both datasets. Thus, our dataset does not contain browsing behaviors (w_i) for users who *never* visited the advertiser's website during the observation window. Consider a set of non-visitors N and a set of visitors, E. While the probability that visitor *i* for $i \in E$ belongs to class *k* can be estimated using the equation for m_{ik}, one must make an inference of the membership probability for non-visitors $i'$for $i' \in N$ because of the missing w_i. Because there is no evidence to suggest that *general* online browsing propensities (e.g., online usage frequency, interests in finance, e-commerce, and media, etc.) differ between visitors and non-visitors, the weighted average membership probabilities for each class, estimated from visitors, are applied to non-visitors as well (note that this does not impose equal proportional apportioning into latent classes for the non-visitors vs. visitors). In other words, the estimated coefficient of the constant term $\hat{\gamma}_{0k}$ for visitors, $i \in E$, $\hat{\gamma}_{0k}$, is used for the baseline probabilities for each segment for $i' \in N$:

$$m_{i'k} = \widehat{m}_k = \frac{\exp(\hat{\gamma}_{0k})}{\sum_{j=1}^{K} \exp(\hat{\gamma}_{0j})}.$$

W-C. MODEL ESTIMATION

i. **Metropolis-Hastings (MH) algorithm to update $\boldsymbol{\gamma}$: : $\boldsymbol{\gamma}|\mathbf{C}, \boldsymbol{\mu}_{\gamma}, \boldsymbol{V}_{\gamma}$**

Assuming that $\gamma_1 = 0$ for identification, the full conditional *w*-dimensional vector parameter γ_k*(k* = 1,...,K) is as follows:

$$\pi(\gamma_k | C_i = k, \mu_\gamma, V_\gamma) \propto \prod_{i=1}^{n} m_{ik}^{I(C_i=k)} \pi(\gamma_k) = \prod_{i:C_i=k} \left(\frac{\exp(w_i \gamma_k)}{\sum_{j=1}^{K} \exp(w_i \gamma_j)} \right) N(\gamma_k | \mu_\gamma = 0, V_\gamma = 100 I_w)$$

Since the posterior distribution does not have a closed form, we use a random-walk MH algorithm with a multivariate-t distribution: $\gamma_k^{t+1} = \gamma_k^t + \kappa_\gamma$ where $\kappa_\gamma \sim MVt(0, s_\gamma T_\gamma)$, where T_γ is the empirical covariance from an extended burn-in period (Haario et al. 2005) so that the proposal density follows the approximate posterior covariance. The parameter s_γ is adjusted such that the rejection ratio belongs to .5~.7.

ii. **Updating $\boldsymbol{m_i}$ and $\boldsymbol{C_i}$**

For i = 1,...,N, draw $C_i \in \{1, \dots, K\}$ with probability m_{ik}such that

$$\Pr(C_i = k) = \frac{m_{ik} \prod_{t=1}^{T_i} f(y_{it} | \alpha_k, \lambda_j, u_i) f(u_i | \Sigma_k)}{\sum_{k=1}^{K} m_{ik} \prod_{t=1}^{T_i} f(y_{it} | \alpha_k, \lambda_j, u_i) f(u_i | \Sigma_k)}$$

Missing browsing behavior data for non-visitors. The dataset of our study does not contain browsing behaviors (w_i) for users who never visited the advertiser's website during the observation periods. Let's call a set of non-visitors N and a set of visitors E. Although the probability that visitor i for $i \in$ E belongs to the class k can be estimated by the equation for m, we must make an inference regarding the probability that the non-visitor i' for $i' \in$ N belongs to the class k because of the missing $w_{i'}$. To overcome the missing data problem, we assume that missing browsing behaviors are independent from class allocation, applying baseline probability estimated from visitors to non-visitors (Rubin 1976). In other words, the estimated coefficient of the constant term $\hat{\gamma}_{0k}$ for visitors,$i \in$ E, $\hat{\gamma}_{0k}$, is used for the baseline probabilities for each segment for $i' \in$ N:

$$\mathrm{m}_{i'k} = \widehat{m}_k = \frac{\exp(\hat{\gamma}_{0k})}{\sum_{j=1}^{K} \exp(\hat{\gamma}_{0j})}$$

iii. **MH algorithm to update $\boldsymbol{\alpha_k}$: $\boldsymbol{\alpha_k} | \mathbf{C}_{\boldsymbol{i}} = \boldsymbol{k}, \boldsymbol{\mu_\alpha}, \boldsymbol{\Lambda_\alpha}$**

Let $\mathrm{M_{ik}} = [1 \quad AdStock_i \quad AdStock_i^2 \quad z_i \quad z_i^2 \quad v_i \quad v_i A_i \quad v_i z_i]$ denote the $\mathrm{T_i} \times r$ submatrix of covariates for an user *i* such that $\mathrm{C_i} = k$, α_{1k} denote a $2 \times \mathrm{r}$ coefficient matrix for the multinomial component with the first column corresponding to one visit and the second column to more than one visit, and α_{2k} denote a r coefficient vector for the discretized lognormal component. Note that simultaneous updating α_{1k} and α_{2k} is possible while in our paper, we sequentially update them for the better and more efficient mix. .

$$\pi(\alpha_{1k} | C_i = k, \mu_{\alpha 1}, \Lambda_{\alpha 1}) \propto \prod_{i:C_i=k} \left[\Pr(y_i = q | \alpha_{1k}, u_{iq})\right]^{I(y_i=q)} \pi(\alpha_{1k})$$

$$= \prod_{i:C_i=k} \prod_{t=1}^{T_i} \prod_{q=0}^{2} \left[\frac{\exp(M_{ik}\alpha_{kq} + u_{iq})}{\sum_{r=1}^{2} \exp(M_{ik}\alpha_{kr} + u_{ir})}\right]^{I(y_i=q)} N(\alpha_{1k} | \mu_{\alpha 1} = 0, \Lambda_{\alpha 1}$$

$$= 100 I_{\alpha 1})$$

$$\pi(\alpha_{2k} | C_i = k, \mu_{\alpha 2}, \Lambda_{\alpha 2}, \boldsymbol{\tau}_\mathbf{k}) \propto \prod_{i:C_i=k} [\Pr(y_i = q | \alpha_{2k}, u_{i3}, \boldsymbol{\tau}_\mathbf{k})]^{I(y_i=q)} \pi(\alpha_{2k})$$

$$= \prod_{i:C_i=k} \prod_{t=1}^{T_i} \prod_{q=0}^{2} p_2(Y = y_{it} | \mu_{it}, \tau_k)^{I(y_{it} \geq 2)} N(\alpha_{2k} | \mu_{\alpha 2} = 0, \Lambda_{\alpha 2} = 100 I_{\alpha 2})$$

Since the posterior distribution does not have a closed form, we use a random-walk MH algorithm with a multivariate-*t* jumping distribution:

$$\alpha_k^{(t+1)} = \alpha_k^{(t)} + \kappa_{\alpha k} \text{ where } \kappa_{\alpha k} \sim MVt(0, s_{\alpha k} T_{\alpha k}) \text{ for both } \alpha_{1k} \text{ and } \alpha_{2k},$$

where parameters $\mathrm{T}_{\alpha k}$ and $\mathrm{s}_{\alpha k}$ are set up as in the estimation of γ_k.

iv. **MH algorithm to update $\boldsymbol{\tau_k}$**

As τ_k is defined in a positive domain, we use the lognormal distribution $LN(\tau_k | \mu_\tau = 0, \sigma_\tau = 0.1)$ for its prior distribution, indicating its posterior distribution as below.

$$\pi(\tau_k | C_i = k, \mu_{\alpha 2}, \Lambda_{\alpha 2}, \alpha_{2k}) \propto \prod_{i:C_i=k} [\Pr(y_i = q | \alpha_{2k}, u_{i3}, \tau_k)]^{I(y_i=q)} \pi(\tau_k)$$

$$= \prod_{i:C_i=k} \prod_{t=1}^{T_i} \prod_{q=0}^{2} p_2(Y = y_{it} | \mu_{it}, \tau_k)^{I(y_{it} \geq 2)} LN(\tau_k | \mu_\tau = 0, \sigma_\tau = 0.1)$$

For proposal density, we use a random-walk algorithm with an exponential normal distribution.

$$\tau_k^{(t+1)} = \exp(\tau_k^{(t+1)} + o_k) \text{ where } o_k \sim N(0, 0.1)$$

v. **MH algorithm to draw $\boldsymbol{u_{ik}}$**

$$\pi(u_{ik}|C_i = k, y_{it}, \alpha_k, \tau_k, AdStock_{it}) \propto f(y_{it}|C_i = k, \alpha_k, u_{ik}, \tau_k, AdStock_{it}) N(u_{ik}|0, \Sigma_k)$$

For proposal density, we use a random-walk MH algorithm with a multivariate normal distribution with variance Σ_k:

$$u_{ik}^{(t+1)} = u_{ik}^{(t+1)} + \kappa_{uk} \text{ where } \kappa_{uk} \sim MVN(0, \Sigma_k)$$

vi. **Gibbs sampling to update $\boldsymbol{\Sigma_k}$**

$$\pi(\Sigma_k|u_k) \sim IW\left(n_k + \nu_0, D_0 + u_k^{'} u_k\right)$$

vii. **MH algorithm to update the website random effects $\boldsymbol{\lambda_j}$**

The posterior distribution of λ_j (j = 1, ..., J) is as follows

$$\pi(\lambda_j|X, y, \alpha_k, \beta_k, u_k, \tau_k^2, m_k, \theta)$$
$$\propto \prod_{i=1}^{N} \left\{ \sum_{k=1}^{K} f(m_{ik}|\gamma_{ik}, w_i) \left[\prod_{t=1}^{T_i} f\big(y_{it}\big|\alpha_k, \lambda_j, u_i, \tau_k\big) f(u_i|\Sigma_k) \right] \right\} \prod_{j=1}^{J} \pi(\lambda_j|\theta)$$

where the prior distribution $\pi\big(\lambda_j\big|\theta\big) = \frac{1}{\Gamma\left(\frac{1}{\theta}\right)\theta^{1/\theta}} \lambda_j^{\frac{1}{\theta}-1} \exp\left(-\frac{\lambda_j}{\theta}\right)$ follows the gamma distribution with a mean of one and a variance of θ.

For the proposal density, we use an independence sampler with gamma distribution (Hahn, 2014), $g(\lambda_j^{(t+1)}|\lambda_j^{(t)}) \sim Gamma(\iota, \frac{\lambda_j^{(t)}}{\iota})$ where ι is the shape parameter and $\frac{\lambda_j^{(t)}}{\iota}$ is the scale parameter.

viii. **MH algorithm to update $\boldsymbol{\theta}$**

A prior density for θ is assumed as a uniform distribution, Unif$(0, \infty)$, and the independent sampler with gamma distribution is used for the proposal density.

[REFERENCES]

Rubin, D. B. 1976. Inference and missing data. *Biometrika*, 63 581-592.

Haario, H., E.Saksman, J. Tamminen. 2005. Componentwise adaptation for high dimensional MCMC. *Computational Statistics* 20(2), 265–273.

W-D. ROBUSTNESS CHECKS

To ensure that claims regarding weariness do not hinge on particular modeling assumptions, we examine variants of the proposed model that turn off or hobble specific parts of it. Our initial set of checks focus on the lognormal portion of the count model. We first examine a binary model that distinguishes visits as 0 vs. 1-or-more; that is, the proposed model with the multinomial portion made binary, and with no discretized lognormal component. Results (Web Appendix, Table W-6 and Figure W-1) indicate that, for this model, four latent classes fit best. Classes 1 and 4 show mean contours consistent with response concavity (weariness), although, due to information loss (i.e., the number of visits), neither rises to the 95% contour concavity standard apparent for the full model. We note in passing that it is precisely this information – the number of individual visits – that is differentially important in determining eventual conversion, as pointed out by Moe and Fader (2004). Using the binary incidence DV precludes fit statistics from being compared to those from our best-fitting model. We can, however, re-estimate our model with no covariates (i.e., just intercepts) in the discretized lognormal model, meaning a multinomial model separating 0 vs. 1 vs. 2+, which is nested in our model. Model fit statistics (Table W-7, for 3-, 4-, and the best-fitting 5-class model) show that removing covariates from the discretized lognormal portion (i.e., predicting distribution *among* the 2+ visit observations) produces greatly inferior results, roughly 10% worse on LMD, DIC, elpd.waic, and elpd.isloo.

We further checked whether the existence of weary groups hinges on particular model specifications (beyond the lognormal portion) or sample selection criteria; we do this in six ways. First, we find a distinct weary class even when imposing equal publisher effectiveness ($\lambda_j = 1$; Figure W-2). Second, we re-estimate the model using the same AdStock specification (Eq. (7)) but with *SAI* in place of log(1 + *SAI*); this results (Figure W-3) in five classes, one of which displays strong weariness, but a dramatically poorer overall fit (roughly 15% in each of the four fit metrics). Third, we excluded outliers using different percentiles (i.e., 0.1%, 1%, 5%, vs. the 0.01%

previously); in each case, there was at least one class with size greater than 30% and over 95% weary contours. The last set of checks concerns the potential for so-called "activity bias" (Lewis and Reiley 2014), where very active internet users are more likely to both receive a high number exposures and respond differently to advertising (compared with lighter users). This concern is addressed in two distinct ways: (a) we constructed a daily browsing activity variable (i.e., the number of *daily* webpages visited), using it as a control covariate; and (b) constructed two additional sample datasets of 12,000 users based on overall browsing activities, specifically, the 2nd and 4th quintiles of users' *total* number of webpages visited). As before, models run in all these cases perfectly replicate our substantive results: a single class with significant, substantial weariness (Figure W-4).[1] In short, weariness was robust to all critical sample selection criteria and model constructs, save one: response *homo*geneity, which we turn to next.

[1] Detailed estimation results for all these models are available from the authors.

Table W-1: Simulation Study Based on the Actual Data Set

(Burn-in 5,000/10,000 draws)

Parameters	True value	Median	2.5%	97.5%		Parameters	True value	Median	2.5%	97.5%
$\alpha_{k=1,q=1}$	-1.21	-1.12	-1.37	-0.92		$\alpha_{k=1,q=3}$	-0.071	-0.053	-0.071	-0.013
	-1.11	-1.06	-1.2	-0.92			2.176	2.177	2.17	2.182
	-0.8	-0.67	-0.8	-0.55			2.483	2.483	2.474	2.49
	1.24	1.17	0.96	1.41			0.92	0.877	0.822	0.91
	0.32	0.37	0.14	0.57			-0.226	-0.205	-0.22	-0.181
	-0.92	-0.85	-1.07	-0.6			0.71	0.694	0.672	0.71
$\alpha_{k=1,q=2}$	-1.04	-1.05	-1.19	-0.93		$\alpha_{k=2,q=3}$	-1.036	-1.037	-1.043	-1.031
	2.05	2.14	2	2.28			-0.936	-0.933	-0.94	-0.927
	2.2	2.03	1.72	2.27			-0.074	-0.087	-0.099	-0.073
	1.19	1.25	1.03	1.48			0.396	0.378	0.353	0.394
	-1.52	-1.35	-1.72	-1.01			-0.504	-0.541	-0.565	-0.517
	1.7	1.75	1.57	1.88			-1.986	-1.953	-1.977	-1.93
$\alpha_{k=2,q=1}$	-2.29	-2.21	-2.35	-2.07		$\alpha_{k=3,q=3}$	-0.204	-0.243	-0.261	-0.225
	-1.54	-1.37	-1.67	-1.03			-0.631	-0.588	-0.614	-0.563
	-5.05	-4.83	-5.24	-4.45			0.535	0.493	0.469	0.516
	0.83	0.74	0.6	0.88			-0.071	-0.053	-0.071	-0.013
	-2.16	-2.08	-2.19	-1.98			2.176	2.177	2.17	2.182
	-1.32	-1.29	-1.38	-1.18			2.483	2.483	2.474	2.49
$\alpha_{k=2,q=2}$	0.23	0.31	0.14	0.48						
	-0.87	-0.86	-0.95	-0.78		$\tau_{k=1}$	0.01	0.018	0.017	0.021
	-1.23	-1.26	-1.65	-0.92		$\tau_{k=2}$	0.01	0.019	0.018	0.021
	1.6	1.45	1.32	1.59		$\tau_{k=3}$	0.01	0.101	0.097	0.107
	1.33	1.28	1.15	1.46						
	-1.57	-1.4	-1.7	-1.09		$\gamma_{k=2}$	-1.3	-1.4	-1.8	-1.2
$\alpha_{k=3,q=1}$	-3.05	-2.98	-3.44	-2.55			-4	-4.4	-5	-3.9
	1.18	1.12	0.92	1.28			1.8	2	1.6	2.3
	-0.21	-0.25	-0.36	-0.14		$\gamma_{k=3}$	-1.5	-1.1	-1.4	-0.9
	-1.55	-1.55	-1.66	-1.42			2.9	2.4	2	2.7
	0.34	0.34	0.17	0.53			-1.7	-1.6	-1.8	-1.3
	0.34	0.32	0.21	0.43						
$\alpha_{k=3,q=2}$	-1.21	-1.12	-1.37	-0.92		θ	1	0.69	1.21	0.42
	-1.11	-1.06	-1.2	-0.92						
	-0.8	-0.67	-0.8	-0.55		$\rho(\lambda,\tilde{\lambda})$	0.999			
	1.24	1.17	0.96	1.41						
	0.32	0.37	0.14	0.57						
	-0.92	-0.85	-1.07	-0.6						

TABLE W-2. T-tests for the Number of Impressions Before vs. After visits

	Total number of impressions before, and after, the first visit	Number of impressions one week before, and after, the first visit	Number of impressions 6-8 days before, and after, the first visit
Obs.	4,135	4,135	4,135
MEAN Before	3.16	1.21	0.34
MEAN After	1.95	0.88	0.22
MEAN Diff	1.21	0.33	0.13
Std. Err	0.08	0.27	0.02
Std. Dev	5.26	1.73	1.03
95% CI – lower	1.05	0.28	0.96
– upper	1.37	0.38	0.16
p-value*	0.000	0.000	0.000

**H_0*: The total number of impressions before and after the first visit is the same (i.e., difference =0)*

TABLE W-3. Endogeneity Tests for Stratified Number of Impressions

Control	Treatment	Obs.	Mean	Std. Err	95% CI		p-value	.05 Crit. Diff
week 2:	week 3: no visit	1,307	0.33	0.03	0.27	0.39	0.14	0.12
1 imp.	week 3: first visit	266	0.44	0.08	0.28	0.6		
week 2:	week 3: no visit	364	0.61	0.07	0.46	0.75	0.02	0.23
2 imp.	week 3: first visit	40	1.18	0.25	0.67	1.68		
week 2:	week 3: no visit	59	3.2	0.39	2.42	3.98	0.27	0.22
3 imp.	week 3: first visit	9	2	1.07	-0.47	4.46		
week 2:	week 3: no visit	29	2.59	0.6	1.36	3.81	0.63	0.24
5 imp.	week 3: first visit	3	1.67	1.2	-3.5	6.84		
week 3:	week 4: no visit	1,815	0.38	0.02	0.34	0.42	0.91	0.40
1 imp.	week 4: first visit	306	0.39	0.05	0.29	0.48		
week 3:	week 4: no visit	537	0.62	0.05	0.52	0.73	0.35	0.13
2 imp.	week 4: first visit	70	0.77	0.16	0.44	1.1		
week 3:	week 4: no visit	204	1.08	0.13	0.83	1.33	0.37	0.20
3 imp.	week 4: first visit	31	1.39	0.29	0.79	1.98		
week 3:	week 4: no visit	94	1.94	0.22	1.5	2.38	0.72	0.21
4 imp.	week 4: first visit	11	2.18	0.62	0.81	3.55		
week 3:	week 4: no visit	45	2.04	0.3	1.43	2.66	0.76	0.30
5 imp.	week 4: first visit	12	2.25	0.64	0.84	3.66		
week 4:	week 5: no visit	1,829	0.24	0.01	0.21	0.27	0.51	0.31
1 imp.	week 5: first visit	355	0.26	0.04	0.18	0.34		
week 4:	week 5: no visit	489	0.53	0.06	0.42	0.64	0.98	0.17

2 imp.	week 5: first visit	85	0.53	0.11	0.3	0.75		
week 4:	week 5: no visit	241	0.81	0.08	0.65	0.98	0.32	0.19
3 imp.	week 5: first visit	27	1.07	0.28	0.5	1.64		
week 4:	week 5: no visit	78	1.1	0.16	0.78	1.42	0.02	0.31
4 imp.	week 5: first visit	12	2.25	0.54	1.07	3.43		
week 4:	week 5: no visit	40	1.6	0.27	1.05	2.15	0.42	0.40
5 imp.	week 5: first visit	9	2.11	0.56	0.81	3.41		
week 5:	week 6: no visit	1,951	0.07	0.01	0.04	0.09	0.44	0.19
1 imp.	week 6: first visit	405	0.04	0.01	0.02	0.06		
week 5:	week 6: no visit	676	0.12	0.02	0.08	0.16	0.29	0.36
2 imp.	week 6: first visit	107	0.19	0.08	0.04	0.34		
week 5:	week 6: no visit	279	0.15	0.03	0.09	0.21	0.94	0.57
3 imp.	week 6: first visit	51	0.16	0.06	0.03	0.29		
week 5:	week 6: no visit	121	0.29	0.06	0.17	0.41	0.18	0.59
4 imp.	week 6: first visit	26	0.5	0.19	0.12	0.88		
week 5:	week 6: no visit	70	0.37	0.1	0.18	0.57	0.41	0.78
5 imp.	week 6: first visit	12	0.17	0.17	-0.2	0.53		
week 6:	week 7: no visit	1,738	0.07	0.01	0.05	0.09	0.55	0.26
1 imp.	week 7: first visit	379	0.08	0.03	0.03	0.14		
week 6:	week 7: no visit	537	0.26	0.07	0.12	0.39	0.79	0.15
2 imp.	week 7: first visit	90	0.21	0.06	0.09	0.33		
week 6:	week 7: no visit	189	0.26	0.09	0.08	0.44	0.58	0.34
3 imp.	week 7: first visit	30	0.13	0.06	0.004	0.26		
week 6:	week 7: no visit	90	0.48	0.16	0.15	0.8	0.83	0.40
4 imp.	week 7: first visit	13	0.38	0.18	-0.01	0.78		
week 6:	week 7: no visit	33	0.24	0.12	0.01	0.48	0.98	1.68
5 imp.	week 7: first visit	4	0.25	0.25	-0.55	1.05		

TABLE W-4: MODEL FIT COMPARISONS

Model						epld	
Carryover (δ)	Classes	$\{\lambda_j\}$	$E(\theta)$	LMD	DIC	WAIC	IS-LOO
10%	K=1	Yes	0.50	-166020	329336	-176860	-180213
				-166021	328608	-177016	-180418
	K=3	Yes	0.87	-152522	286696	-163433	-166621
				-152508	289103	-163920	-166805
20%	K=1	Yes	0.23	-165682	326851	-176615	-179916
				-165718	327776	-176832	-180235
	K=2	Yes	0.49	-141898	271173	-150154	-153956
				-142362	276803	-152050	-156112
30%	K=1	Yes	0.96	-165637	327382	-176564	-179893
				-165742	325415	-176560	-179904
	K=4	Yes	0.79	-152144	293936	-165854	-168878
				-159035	312103	-168977	-172395
40%	K=1	Yes	1.03	-165662	327210	-176630	-179990
				-165975	328088	-176684	-180062
	K=4	Yes	0.52	-153631	301742	-165682	-168509
				-154340	296383	-166605	-169238
50%	K=1	Yes	0.23	-165953	325984	-176606	-179984
				-165774	326635	-176652	-180039
	K=5	Yes	0.24	-141186	270512	-152597	-156301
				-143223	276504	-152508	-156065
60%	K=1	Yes	0.23	-165953	325984	-176606	-179984
				-165774	326635	-176652	-180039
	K=4	Yes	0.24	-140959	276843	-154910	-161630
				-142632	271013	-158048	-163972
	K=5	Yes	0.25	-139589	269451	-150462	-155386
				-141729	278867	-152785	-156763
	K=6	Yes	0.24	-142777	277176	-154219	-159586
				-143167	274736	-156607	-163828
70%	K=1	Yes	0.31	-166015	327293	-176675	-180035
				-166437	329621	-178225	-182604
	K=5	Yes	0.25	-143909	267328	-159840	-164406
				-148332	287013	-162173	-165889
80%	K=1	Yes	0.30	-166342	329918	-177175	-180596
				-166260	330334	-177294	-180653
	K=5	Yes	0.25	-146998	279887	-163584	-166509
				-148868	284177	-163717	-166146
90%	K=1	Yes	0.53	-166077	330334	-176947	-180343
				-166393	328640	-177086	-180458
	K=5	Yes	0.44	-143909	267328	-159840	-164406
				-148332	287013	-162173	-165889

Note: Model fit metrics for various models using different carryover parameters (δ), number of classes (K), with and without publisher heterogeneity.

Table W-5: SIZE OF LATENT CLASSES FOR 60% CARRYOVER MODEL

	$\boldsymbol{E(\theta)}$	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**
$K = 1$	0.23	100.00%					
$K = 2$	0.29	34.23%	65.77%				
$K = 3$	0.27	33.03%	59.27%	7.70%			
$K = 4$	0.24	3.53%	60.83%	7.78%	27.88%		
$K = 5$	0.25	2.35%	24.37%	7.80%	57.78%	7.71%	
$K = 6$	0.24	0.12%	13.11%	7.82%	17.21%	7.31%	54.44%

Table W-6: FIT STATISTICS FOR BINARY MODEL

	LMD	DIC	elpd.waic	elpd.isloo
3 class	-94605	199437	-98495	-100557
4 class*	-88365	185956	-91475	-93535
5 class	-91386	189967	-93973	-96706

* Latent class sizes are 23%, 14%, 17%, and 46%

Table W-7: MODEL WITH NO COVARIATES IN DISCRETIZED LOGNORMAL

	LMD	DIC	elpd.waic	elpd.isloo
3 class	-152670	296337	-164486	-167314
4 class	-152711	297088	-164403	-167598
5 class	-152225	296166	-163302	-165742
Proposed*	-139589	269451	-150462	-155386

* Proposed model with 60% carryover and 5 Classes

FIGURE W-1: RESPONSE SHAPES FOR BINARY MODEL

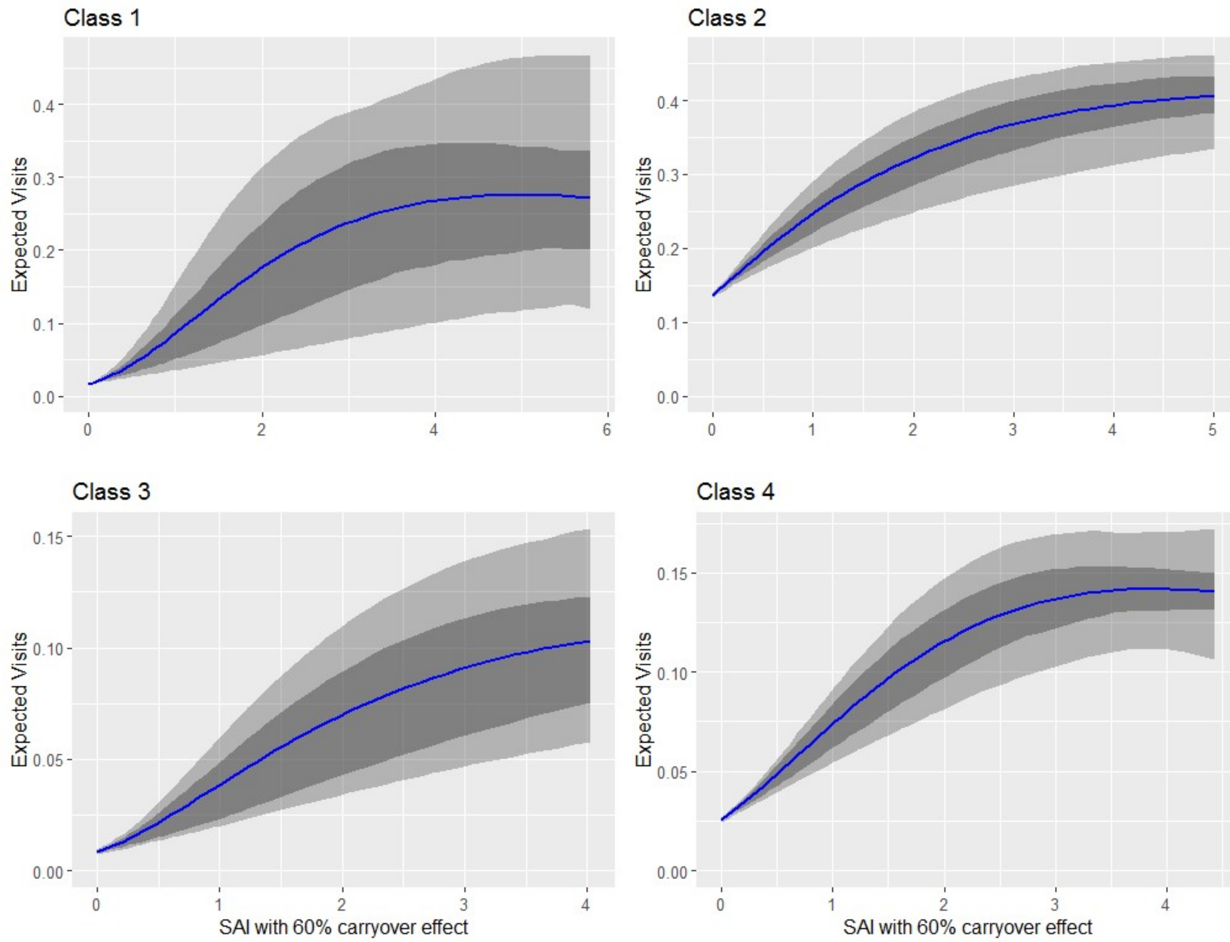

FIGURE W-2: RESPONSE SHAPES WITH EQUAL PUBLISHER EFFECTIVENESS

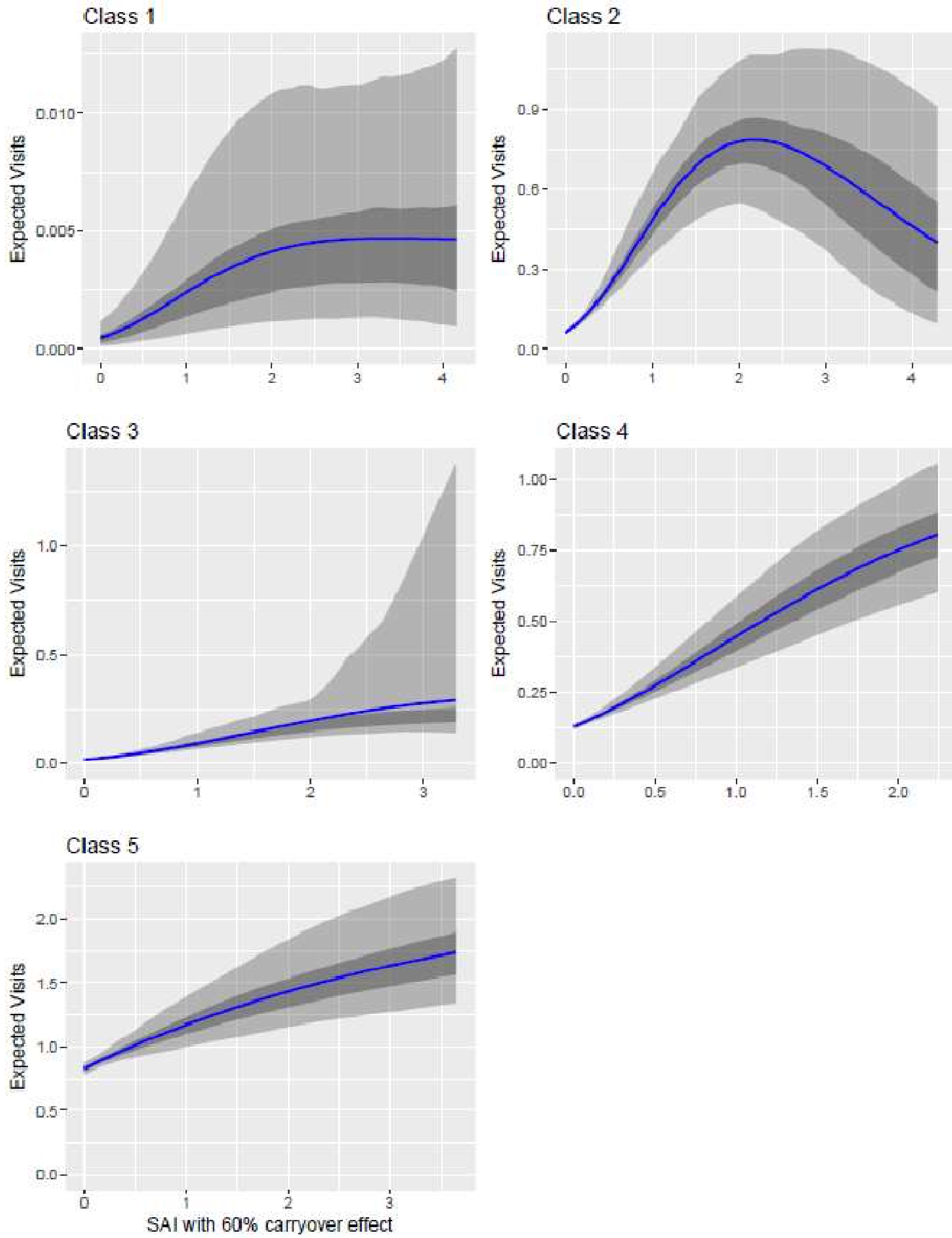

FIGURE W-3: RESPONSE SHAPESOF THE NON-LOG-TRANSFORMED MODEL

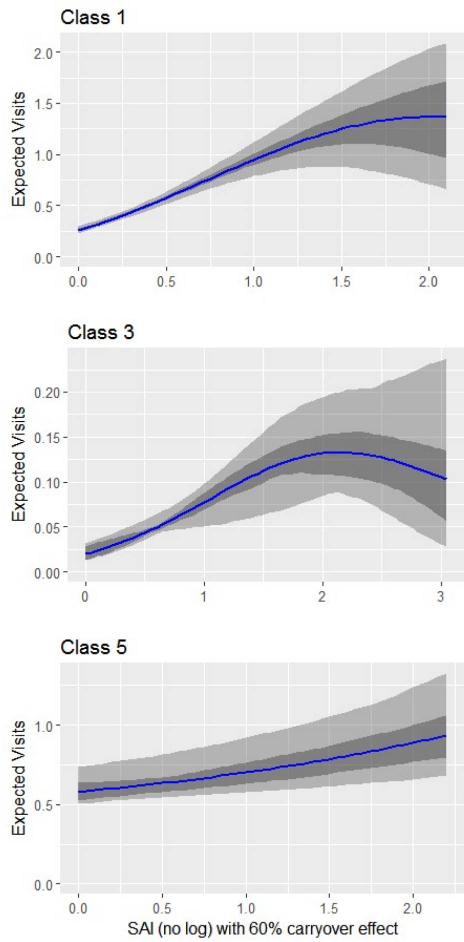

FIGURE W-4: RESPONSE SHAPES OF WEARY CLASSES FOR MODELS

Browsing activity variable (left), less-active users (middle), more-active users (right)