Print Date: 2/23/2004

Technical Appendices:

Modeling Online Browsing and Path Analysis Using Clickstream Data

Alan L. Montgomery, Shibo Li, Kannan Srinivasan, and John C. Liechty

February 2004

Alan L. Montgomery (e-mail: alan.montgomery@cmu.edu) is an Associate Professor at Graduate School of Industrial Administration, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. Shibo Li (shibo_li@rbsmail.rutgers.edu) is an Assistant Professor of Marketing at Rutgers University, 228 Janice Levin Building, 94 Rockafeller Road, Piscataway, NJ 08854. Kannan Srinivasan (kannans@andrew.cmu.edu) is H.J. Heinz II Professor of Management, Marketing, and Information Systems and Director of the Center for E-Business Innovation at the Graduate School of Industrial Administration, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. John C. Liechty (jcl12@psu.edu) is an Assistant Professor of Marketing and Statistics at the Pennsylvania State University, 710 M Business Administration Building, University Park, PA 16802.

Appendix A: Prior Settings

We assume *a priori* that the densities of the starting values for $\boldsymbol{D}_{iq}$, $\boldsymbol{v}_i$, and the density of each row of the hidden chain transition matrix are Dirichlet densities. In addition, we assume *a priori* that the intensity parameters, λ_{is}, for $\boldsymbol{D}_{iq}$ follow a Gamma density. For notational simplicity we drop the dependence of state on the user, session, and viewing (<*iqt*>) and assume that this dependence is understood. That is,

$$\boldsymbol{v}_i \sim Dirichlet(\boldsymbol{\alpha})$$
$$\lambda_{is} \sim Gamma(shape_{prior}, scale_{prior})$$
$$\boldsymbol{P}_{ij} \sim Dirichlet[\boldsymbol{\tau}_j], \boldsymbol{\tau}_j \text{ is a } 1\times \text{S vector, } \boldsymbol{P}_{ij} \text{ is the j}^{\text{th}} \text{ row of } \boldsymbol{P}_i$$
$$\alpha_s \sim Gamma(shape_\alpha, scale_\alpha)$$
$$\tau_{js} \sim Gamma(shape_\tau, scale_\tau)$$
$$shape_{prior} \sim Gamma(shape_\lambda, scale_\lambda)$$
$$scale_{prior} \sim Gamma(shape_\lambda, scale_\lambda)$$

Notice that in the paper we use $\widehat{\lambda}_{sh}$ and $\breve{\lambda}_{sc}$ respectively to denote $shape_{prior}$ and $scale_{prior}$.

To complete our model we specify the following priors:

$$\boldsymbol{\Pi}_{ls} \sim MVN[\overline{\boldsymbol{\Pi}}_s, \boldsymbol{V}_{\boldsymbol{\Pi}}], \text{ for } l \text{ and } s.$$
$$\overline{\boldsymbol{\Pi}}_s \sim MVN[\overline{\overline{\boldsymbol{\Pi}}}, \boldsymbol{V}_{\overline{\boldsymbol{\Pi}}}], \text{ for } s.$$
$$\boldsymbol{V}_{\boldsymbol{\pi}}^{-1} \sim Wishart[\rho_\Pi, \boldsymbol{V}_V].$$
$$\boldsymbol{\Psi}_s^{-1} \sim Wishart[\rho_\varsigma, \boldsymbol{V}_\varsigma], \text{for } s.$$
$$\boldsymbol{\Sigma}_s^{-1} \sim Wishart[\rho_\Sigma, \boldsymbol{V}_{\boldsymbol{\Sigma}}], \text{for } s.$$
$$vec(\overline{\boldsymbol{\Phi}}_s) \sim MVN[\overline{\overline{\boldsymbol{\Phi}}}_s, \boldsymbol{V}_{\overline{\boldsymbol{\Phi}}}]$$
$$\boldsymbol{\Omega}_s^{-1} \sim Wishart[\rho_{\boldsymbol{\Omega}}, \boldsymbol{V}_{\boldsymbol{\Omega}}], \text{for } s.$$

Priors for the Full Models at Various Levels

We assume *a priori* that the densities of the starting values for $\boldsymbol{D}_{iq}$, $\boldsymbol{v}_i$, and the density of each row of the hidden chain transition matrix are Dirichlet densities. In addition, we assume *a priori* that the intensity parameters, λ_{is}, for $\boldsymbol{D}_{iq}$ follow a Gamma density. That is,

$$\boldsymbol{v}_i \sim Dirichlet(\boldsymbol{\alpha})$$
$$\lambda_{is} \sim Gamma(shape_{prior}, scale_{prior})$$
$$\boldsymbol{P}_{ij} \sim Dirichlet[\boldsymbol{\tau}_j], \boldsymbol{\tau}_j \text{ is a } 1 \times \text{S vector}, \boldsymbol{P}_{ij} \text{ is the j}^{\text{th}} \text{ row of } \boldsymbol{P}_i$$
$$\alpha_s \sim Gamma(shape_\alpha, scale_\alpha), shape_\alpha = I/2, scale_\alpha = I/2$$
$$\tau_{js} \sim Gamma(shape_\tau, scale_\tau), shape_\tau = I/2, scale_\tau = I/2$$
$$shape_{prior} \sim Gamma(shape_\lambda, scale_\lambda), shape_\lambda = I/2, scale_\lambda = I/2$$
$$scale_{prior} \sim Gamma(shape_\lambda, scale_\lambda), shape_\lambda = I/2, scale_\lambda = I/2$$

To complete our model we specify the following priors:

$$\boldsymbol{\Pi}_{ls} \sim MVN[\overline{\boldsymbol{\Pi}}_s, \boldsymbol{V}_{\boldsymbol{\Pi}}], \text{ for } l \text{ and } s.$$
$$\overline{\boldsymbol{\Pi}}_s \sim MVN[\overline{\overline{\boldsymbol{\Pi}}}, \boldsymbol{V}_{\overline{\boldsymbol{\Pi}}}], \text{ for } s, \overline{\overline{\boldsymbol{\Pi}}} = \boldsymbol{0}, \boldsymbol{V}_{\overline{\boldsymbol{\Pi}}} = \frac{C+2}{100} \cdot \boldsymbol{I}$$
$$\boldsymbol{V}_{\boldsymbol{\pi}}^{-1} \sim Wishart[\rho_\Pi, \boldsymbol{V}_V], \rho_\Pi = C+2, \boldsymbol{V}_V = \frac{1000}{C+2} \cdot \boldsymbol{I}$$
$$\boldsymbol{\Psi}_s^{-1} \sim Wishart[\rho_\varsigma, \boldsymbol{V}_\varsigma], \text{for } s, \rho_\varsigma = C+2, \boldsymbol{V}_\varsigma = \frac{100}{C+2} \cdot \boldsymbol{I} \quad ,$$
$$\boldsymbol{\Sigma}_s^{-1} \sim Wishart[\rho_\Sigma, \boldsymbol{V}_{\boldsymbol{\Sigma}}], \text{for } s, \rho_\Sigma = C+2, \boldsymbol{V}_{\boldsymbol{\Sigma}} = \frac{1000}{C+2} \cdot \boldsymbol{I}$$
$$vec(\overline{\boldsymbol{\Phi}}_s) \sim MVN[\overline{\overline{\boldsymbol{\Phi}}}_s, \boldsymbol{V}_{\overline{\boldsymbol{\Phi}}}], \overline{\overline{\boldsymbol{\Phi}}}_s = \boldsymbol{0}, \boldsymbol{V}_{\overline{\boldsymbol{\Phi}}} = 1000 \cdot \boldsymbol{I}$$
$$\boldsymbol{\Omega}_s^{-1} \sim Wishart[\rho_{\boldsymbol{\Omega}}, \boldsymbol{V}_{\boldsymbol{\Omega}}], \text{for } s, \rho_{\boldsymbol{\Omega}} = C+2, \boldsymbol{V}_{\boldsymbol{\Omega}} = \frac{100}{C+2} \cdot \boldsymbol{I}$$

Priors for the Latent Class Models at Various Levels

The prior specifications for the latent class models are the same as those for the full models except for the priors for the latent mixture process. Therefore, we only show the priors for the latent mixture process below.

$$p_i^{'} \sim Dirichlet(\boldsymbol{\alpha})$$

$$\alpha_s \sim Gamma(shape_\alpha, scale_\alpha), shape_\alpha = I/2, scale_\alpha = I/2$$

Priors for the Latent Class Models with Tight Priors

In order to approximate the traditional latent class model, we estimate the user-level latent class models with tight priors on the mixture process, $\boldsymbol{\Psi}_s$, and $\boldsymbol{\Omega}_s$. Therefore we have:

$$p_i^{'} \sim Dirichlet(\boldsymbol{\alpha})$$

$$\alpha_s \sim Gamma(shape_\alpha, scale_\alpha), shape_\alpha = I \cdot S, scale_\alpha = I \cdot S$$

$$\boldsymbol{\Psi}_s^{-1} \sim Wishart[\rho_\varsigma, \boldsymbol{V}_\varsigma], \text{for } s, \rho_\varsigma = C+2, \boldsymbol{V}_\varsigma = \frac{1}{C+2} \cdot \boldsymbol{I}$$

$$\boldsymbol{\Omega}_s^{-1} \sim Wishart[\rho_{\boldsymbol{\Omega}}, \boldsymbol{V}_{\boldsymbol{\Omega}}], \text{for } s, \rho_{\boldsymbol{\Omega}} = C+2, \boldsymbol{V}_{\boldsymbol{\Omega}} = \frac{1}{C+2} \cdot \boldsymbol{I}$$

Other prior specifications remain the same as described above.

Sensitivity of the Model to Prior Specifications

To evaluate the sensitivity of our model to prior settings we consider a range of weak, moderate, and strong priors for the parameter of the hidden markov model as well as those for $\boldsymbol{\Omega}_s$ and $\boldsymbol{\Psi}_s$. The settings of these priors are given below. In our paper we report our model using a moderate prior for both sets of parameters.

		Priors for the Parameters of the Hidden Markov		
		Weak	Moderate	Strong
Priors for $\boldsymbol{\Omega}_s$ and $\boldsymbol{\Psi}_s$	Weak	$\begin{cases} \rho_{\varsigma} = C+2 \\ \boldsymbol{V}_{\varsigma} = \frac{1000}{C+2} \cdot \boldsymbol{I} \\ \rho_{\boldsymbol{\Omega}} = C+2 \\ \boldsymbol{V}_{\boldsymbol{\Omega}} = \frac{1000}{C+2} \cdot \boldsymbol{I} \end{cases}$ $\begin{cases} shape_{\alpha} = 1 \\ scale_{\alpha} = 1 \\ shape_{\tau} = 1 \\ scale_{\tau} = 1 \\ shape_{\lambda} = 1 \\ scale_{\lambda} = 1 \\ shape_{\lambda} = 1 \\ scale_{\lambda} = 1 \end{cases}$	$\begin{cases} \rho_{\varsigma} = C+2 \\ \boldsymbol{V}_{\varsigma} = \frac{1000}{C+2} \cdot \boldsymbol{I} \\ \rho_{\boldsymbol{\Omega}} = C+2 \\ \boldsymbol{V}_{\boldsymbol{\Omega}} = \frac{1000}{C+2} \cdot \boldsymbol{I} \end{cases}$ $\begin{cases} shape_{\alpha} = I/2 \\ scale_{\alpha} = I/2 \\ shape_{\tau} = I/2 \\ scale_{\tau} = I/2 \\ shape_{\lambda} = I/2 \\ scale_{\lambda} = I/2 \\ shape_{\lambda} = I/2 \\ scale_{\lambda} = I/2 \end{cases}$	$\begin{cases} \rho_{\varsigma} = C+2 \\ \boldsymbol{V}_{\varsigma} = \frac{1000}{C+2} \cdot \boldsymbol{I} \\ \rho_{\boldsymbol{\Omega}} = C+2 \\ \boldsymbol{V}_{\boldsymbol{\Omega}} = \frac{1000}{C+2} \cdot \boldsymbol{I} \end{cases}$ $\begin{cases} shape_{\alpha} = I \\ scale_{\alpha} = I \\ shape_{\tau} = I \\ scale_{\tau} = I \\ shape_{\lambda} = I \\ scale_{\lambda} = I \\ shape_{\lambda} = I \\ scale_{\lambda} = I \end{cases}$

	Moderate	$\begin{cases}\rho_{\varsigma}=C+2\\ \boldsymbol{V}_{\varsigma}=\frac{100}{C+2}\cdot\boldsymbol{I}\\ \rho_{\boldsymbol{\Omega}}=C+2\\ \boldsymbol{V}_{\boldsymbol{\Omega}}=\frac{100}{C+2}\cdot\boldsymbol{I}\end{cases}$ $\begin{cases}shape_{\alpha}=1\\ scale_{\alpha}=1\\ shape_{\tau}=1\\ scale_{\tau}=1\\ shape_{\lambda}=1\\ scale_{\lambda}=1\\ shape_{\lambda}=1\\ scale_{\lambda}=1\end{cases}$	$\begin{cases}\rho_{\varsigma}=C+2\\ \boldsymbol{V}_{\varsigma}=\frac{100}{C+2}\cdot\boldsymbol{I}\\ \rho_{\boldsymbol{\Omega}}=C+2\\ \boldsymbol{V}_{\boldsymbol{\Omega}}=\frac{100}{C+2}\cdot\boldsymbol{I}\end{cases}$ $\begin{cases}shape_{\alpha}=I/2\\ scale_{\alpha}=I/2\\ shape_{\tau}=I/2\\ scale_{\tau}=I/2\\ shape_{\lambda}=I/2\\ scale_{\lambda}=I/2\\ shape_{\lambda}=I/2\\ scale_{\lambda}=I/2\end{cases}$	$\begin{cases}\rho_{\varsigma}=C+2\\ \boldsymbol{V}_{\varsigma}=\frac{100}{C+2}\cdot\boldsymbol{I}\\ \rho_{\boldsymbol{\Omega}}=C+2\\ \boldsymbol{V}_{\boldsymbol{\Omega}}=\frac{100}{C+2}\cdot\boldsymbol{I}\end{cases}$ $\begin{cases}shape_{\alpha}=I\\ scale_{\alpha}=I\\ shape_{\tau}=I\\ scale_{\tau}=I\\ shape_{\lambda}=I\\ scale_{\lambda}=I\\ shape_{\lambda}=I\\ scale_{\lambda}=I\end{cases}$
	Strong	$\begin{cases}\rho_{\varsigma}=C+2\\ \boldsymbol{V}_{\varsigma}=\frac{1}{C+2}\cdot\boldsymbol{I}\\ \rho_{\boldsymbol{\Omega}}=C+2\\ \boldsymbol{V}_{\boldsymbol{\Omega}}=\frac{1}{C+2}\cdot\boldsymbol{I}\end{cases}$ $\begin{cases}shape_{\alpha}=1\\ scale_{\alpha}=1\\ shape_{\tau}=1\\ scale_{\tau}=1\\ shape_{\lambda}=1\\ scale_{\lambda}=1\\ shape_{\lambda}=1\\ scale_{\lambda}=1\end{cases}$	$\begin{cases}\rho_{\varsigma}=C+2\\ \boldsymbol{V}_{\varsigma}=\frac{1}{C+2}\cdot\boldsymbol{I}\\ \rho_{\boldsymbol{\Omega}}=C+2\\ \boldsymbol{V}_{\boldsymbol{\Omega}}=\frac{1}{C+2}\cdot\boldsymbol{I}\end{cases}$ $\begin{cases}shape_{\alpha}=I/2\\ scale_{\alpha}=I/2\\ shape_{\tau}=I/2\\ scale_{\tau}=I/2\\ shape_{\lambda}=I/2\\ scale_{\lambda}=I/2\\ shape_{\lambda}=I/2\\ scale_{\lambda}=I/2\end{cases}$	$\begin{cases}\rho_{\varsigma}=C+2\\ \boldsymbol{V}_{\varsigma}=\frac{1}{C+2}\cdot\boldsymbol{I}\\ \rho_{\boldsymbol{\Omega}}=C+2\\ \boldsymbol{V}_{\boldsymbol{\Omega}}=\frac{1}{C+2}\cdot\boldsymbol{I}\end{cases}$ $\begin{cases}shape_{\alpha}=I\\ scale_{\alpha}=I\\ shape_{\tau}=I\\ scale_{\tau}=I\\ shape_{\lambda}=I\\ scale_{\lambda}=I\\ shape_{\lambda}=I\\ scale_{\lambda}=I\end{cases}$

Appendix B: Monte Carlo Markov Chain for Estimating the Model

We apply data augmentation and MCMC methods (Gibbs sampler see McCulloch and Rossi (1994) and reversible jump algorithm (Liechty and Roberts, 2001)) based on the following full conditional distributions. For notational simplicity we drop the dependence of state on the user, session, and viewing (<*iqt*>) and assume that this dependence is understood.

$$\boldsymbol{U}_{iqct} \mid \{\boldsymbol{U}_{iqc\bar{t}}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{\Gamma}_{is}\}, \{X_{iqt}\}, \boldsymbol{\Sigma}_s, D_{iqt}$$

$$\boldsymbol{\Gamma}_{is} \mid \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, \boldsymbol{R}_i, \boldsymbol{\Pi}_s, \boldsymbol{\Psi}_s, D_{iqt}$$

$$\boldsymbol{\Phi}_{ils} \mid \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{X}_{iqt}\}, \{\boldsymbol{\Gamma}_{is}\}, \boldsymbol{\Sigma}_s, \overline{\boldsymbol{\Phi}}_{ls}, \boldsymbol{\Omega}_S, D_{iqt}$$

$$\boldsymbol{\Pi}_s \mid \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{R}_i\}, \boldsymbol{\Psi}_s, \overline{\boldsymbol{\Pi}}_s, V_{\boldsymbol{\Pi}}$$

$$\boldsymbol{\Psi}_s^{-1} \mid \{\boldsymbol{\Gamma}_{is}\}, \boldsymbol{\Pi}_s, \{\boldsymbol{R}_i\}, \rho_{\varsigma}, \boldsymbol{V}_{\varsigma}$$

$$\boldsymbol{\Sigma}_s^{-1} \mid \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iqt}\}, \rho_{\Sigma}, \boldsymbol{V}_{\Sigma}$$

$$\lambda_{\mathrm{is}} \mid \{\boldsymbol{D}_{iq}\}, shape_{prior}, scale_{prior}$$

$$\boldsymbol{D}_{iq} \mid \{\boldsymbol{U}_{iqct}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, \{\boldsymbol{\Gamma}_{is}\}, \boldsymbol{v}_i, \boldsymbol{P}_{ij}$$

$$\boldsymbol{v}_i \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{\alpha}$$

$$\boldsymbol{P}_{ij} \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{v}_i, \boldsymbol{\tau}_j$$

$$shape_{prior} \mid \{\boldsymbol{D}_{iq}\}, \lambda_{\mathrm{is}}, scale_{prior}, shape_{\lambda}, scale_{\lambda}$$

$$scale_{prior} \mid \{\boldsymbol{D}_{iq}\}, shape_{prior}, \lambda_{\mathrm{ss}}, shape_{\lambda}, scale_{\lambda}$$

$$\alpha_s \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{v}_i, shape_{\alpha}, scale_{\alpha}$$

$$\tau_{sj} \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{P}_{ij}, shape_{\tau}, scale_{\tau}$$

The univariate draws of $\boldsymbol{U}_{iqct}$ can be generated from truncated normal distribution using a Slice Sampler. The multivariate draws of $\boldsymbol{\Gamma}_{is}$, $\boldsymbol{\Phi}_{ils}$, and $\boldsymbol{\Pi}_{ls}$ can be generated from conjugate multivariate normal distributions with the identification conditions satisfied. The random draws of $\boldsymbol{\Sigma}_s^{-1}$ and $\boldsymbol{\Psi}_s^{-1}$ can be generated from conjugate Wishart distributions. The random draws of λ_{is} can be generated from conjugated Gamma distribution. We can generate the draw of $\boldsymbol{D}_{iq}$ using a reversible jump algorithm (Liechty and Roberts 2001). The multivariate draws of $\boldsymbol{v}_i$ and

$\boldsymbol{P}_{ij}$ can be generated from conjugate Dirichlet distributions. $scale_{prior}$ can be generated from conjugate Gamma distribution. $shape_{prior}$, α_s, and τ_{sj} can be generated using Metropolis-Hasting algorithm. We make a total of 10,000 draws with an additional 5,000 draws for burn-in.

The full conditional distributions are given as follows.

(1). $\boldsymbol{U}_{iq0} \mid \{\boldsymbol{U}_{iq0^-}\}, \{\boldsymbol{Y}_{iq0}\}, \{\boldsymbol{\Phi}_{is}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iq}\}, \boldsymbol{\Sigma}_s, D_{iq0}$

$$\begin{cases} \boldsymbol{U}_{iq1} = \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq0} + \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} + \boldsymbol{\varepsilon}_{iq1s} \\ \boldsymbol{U}_{iq2} = \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq1} + \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq2} + \boldsymbol{\varepsilon}_{iq2s} \\ \cdots \\ \boldsymbol{U}_{iqT_{iq}} = \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq(T_{qi}-1)} + \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqT_{iq}} + \boldsymbol{\varepsilon}_{iqT_{iq}s} \end{cases}$$

$$\Rightarrow \begin{pmatrix} \boldsymbol{U}_{iq1} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} \\ \boldsymbol{U}_{iq2} - \boldsymbol{\Phi}_{is}\boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq2} \\ \cdots \\ \boldsymbol{U}_{iqT_{iq}} - \boldsymbol{\Phi}_{is}^{T_{iq}-1}\boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} - \boldsymbol{\Phi}_{is}^{T_{iq}-2}\boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq2} - \cdots - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqT_{iq}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_{is} \\ \boldsymbol{\Phi}_{is}^2 \\ \cdots \\ \boldsymbol{\Phi}_{is}^{T_i} \end{pmatrix} \boldsymbol{U}_{iq0} + \begin{pmatrix} \boldsymbol{\varepsilon}_{iq1s} \\ \boldsymbol{\Phi}_{is}\boldsymbol{\varepsilon}_{iq1s} + \boldsymbol{\varepsilon}_{iq2s} \\ \cdots \\ \boldsymbol{\Phi}_{is}^{T_{iq}-1}\boldsymbol{\varepsilon}_{iq1s} + \cdots + \boldsymbol{\varepsilon}_{iqT_{iq}s} \end{pmatrix}$$

Note $\{U_{iq0-}\}$ denotes all values of $\{U_{iqt}\}$ other than U_{iq0}.

Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{U}_{iq1} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} \\ \boldsymbol{U}_{iq2} - \boldsymbol{\Phi}_{is}\boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq2} \\ \cdots \\ \boldsymbol{U}_{iqT_{iq}} - \boldsymbol{\Phi}_{is}^{T_{iq}-1}\boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq1} - \boldsymbol{\Phi}_{is}^{T_{iq}-2}\boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq2} - \cdots - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqT_{iq}} \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} \mathbf{\Phi}_{is} \\ \mathbf{\Phi}_{is}^2 \\ \cdots \\ \mathbf{\Phi}_{is}^{T_i} \end{pmatrix}$, and

$\tilde{\boldsymbol{\varepsilon}}_i = \begin{pmatrix} \boldsymbol{\varepsilon}_{iq1s} \\ \boldsymbol{\Phi}_{is}\boldsymbol{\varepsilon}_{iq1s} + \boldsymbol{\varepsilon}_{iq2s} \\ \cdots \\ \boldsymbol{\Phi}_{is}^{T_{iq}-1}\boldsymbol{\varepsilon}_{iq1s} + \cdots + \boldsymbol{\varepsilon}_{iqT_{iq}s} \end{pmatrix}$. Then $\tilde{\boldsymbol{\varepsilon}}_i \sim MVN(\mathbf{0},\ \mathbf{V}_i)$. Let $\mathbf{V}_i^{-1} = \mathbf{ee'}$. We have:

$$\boldsymbol{e}'\boldsymbol{A} = \boldsymbol{e}'\boldsymbol{B}\boldsymbol{U}_{iq0} + \boldsymbol{e}'\tilde{\boldsymbol{\varepsilon}}$$

$$\Rightarrow \boldsymbol{U}_{iq0} \mid \{\boldsymbol{U}_{iq0^-}\}, \{\boldsymbol{Y}_{iq0}\}, \{\boldsymbol{\Phi}_{is}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iq}\}, \boldsymbol{\Sigma}_s, D_{iq0}$$

$$\sim \begin{cases} N_{(\max(0, \boldsymbol{U}_{iq(-c)0}),\ +\infty)}((\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}(\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{A}),\ (\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}) \text{ if } Y_{iqc0} = 1 \\ N_{(-\infty,\ \max(0, \boldsymbol{U}_{iq(-c)0}))}((\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}(\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{A}),\ (\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}) \text{ otherwise} \end{cases}$$

Where $s = D_{iq0}$.

(2). $\boldsymbol{U}_{iqt} \mid \{\boldsymbol{U}_{iqt^-}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{\Phi}_{is}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, D_{iqt}$ for $t \neq 0$

$$\begin{cases} \boldsymbol{U}_{iqt} = \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq(t-1)} + \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqt} + \boldsymbol{\varepsilon}_{iqts} \\ \boldsymbol{U}_{iq(t+1)} = \boldsymbol{\Phi}_{is}U_{iqt} + \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq(t+1)} + \boldsymbol{\varepsilon}_{iq(t+1)s} \end{cases}$$

$$\Rightarrow \begin{cases} -\boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq(t-1)} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqt} = -\boldsymbol{U}_{iqt} + \boldsymbol{\varepsilon}_{iqts} \\ \boldsymbol{U}_{iq(t+1)} - \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iqt} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq(t+1)} = \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iqt} + \boldsymbol{\varepsilon}_{iq(t+1)s} \end{cases}$$

$$\Rightarrow \begin{pmatrix} -\boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq(t-1)} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqt} \\ \boldsymbol{U}_{iq(t+1)} - \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iqt} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq(t+1)} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{I} \\ \boldsymbol{\Phi}_{is} \end{pmatrix} \boldsymbol{U}_{iqt} + \begin{pmatrix} \boldsymbol{\varepsilon}_{iqts} \\ \boldsymbol{\varepsilon}_{iq(t+1)s} \end{pmatrix}$$

Let $\boldsymbol{A} = \begin{pmatrix} -\boldsymbol{\Phi}_{is}\boldsymbol{U}_{iq(t-1)} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqt} \\ \boldsymbol{U}_{iq(t+1)} - \boldsymbol{\Phi}_{is}\boldsymbol{U}_{iqt} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iq(t+1)} \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} -\mathbf{I} \\ \mathbf{\Phi}_{is} \end{pmatrix}$, and $\tilde{\boldsymbol{\varepsilon}}_{it} = \begin{pmatrix} \boldsymbol{\varepsilon}_{iqts} \\ \boldsymbol{\varepsilon}_{iq(t+1)s} \end{pmatrix}$. Then

$\tilde{\boldsymbol{\varepsilon}}_{it} \sim MVN(\mathbf{0}, \mathbf{V}_{it})$. Let $\mathbf{V}_{it}^{-1} = \mathbf{ee'}$. We have:

$\boldsymbol{e}'\boldsymbol{A} = \boldsymbol{e}'\boldsymbol{B}\boldsymbol{U}_{iqc1} + \boldsymbol{e}'\tilde{\boldsymbol{\varepsilon}}_{it}$

$\Rightarrow \boldsymbol{U}_{iqt} \mid \{\boldsymbol{U}_{iqt^-}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{\Phi}_{is}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, D_{iqt}$ for $t \neq 0$

$$\sim \begin{cases} N_{(\max(0,\boldsymbol{U}_{qi(-c)0}),\ +\infty)}((\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}(\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{A}),\ (\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}) \text{ if } Y_{iqc1} = 1 \\ N_{(-\infty,\ \max(0,\boldsymbol{U}_{iq(-c)0}))}((\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}(\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{A}),\ (\boldsymbol{B}'\boldsymbol{ee}'\boldsymbol{B})^{-1}) \text{ otherwise} \end{cases}$$

Where $s = D_{iqt}$.

(3).

$\boldsymbol{\Phi}_{ils} \mid \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, \overline{\boldsymbol{\Phi}_{ls}}, \boldsymbol{\Omega}_{ls}, D_{iqt} \sim MVN(\boldsymbol{B}\cdot\boldsymbol{A}, \boldsymbol{B})\cdot$I($W^a_{iqtS} \geq W^a_{iqt(S-1)} \geq \cdots \geq W^a_{iqt1}$ and $W^b_{iqt1} \geq W^b_{iqt2} \geq \cdots \geq W^b_{iqtS}$)

$$\text{Where } \boldsymbol{B} = (\sum_{q=1}^{Q_i}\sum_{t=1}^{T_{iq}} \boldsymbol{U}_{iql(t-1)}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{U}_{iql(t-1)} + \boldsymbol{\Omega}_{ls}^{-1})^{-1}$$

$$\boldsymbol{A} = \sum_{q=1}^{Q_i}\sum_{t=1}^{T_{iq}} \boldsymbol{U}_{iql(t-1)}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{U}^*_{iqt} + \boldsymbol{\Omega}_{ls}^{-1}\overline{\boldsymbol{\Phi}_{ls}}$$

$$\boldsymbol{U}^*_{iqt} = \boldsymbol{U}_{iqt} - \boldsymbol{\Phi}_{i(-l)s}\boldsymbol{U}_{iq(-l)(t-1)} - \boldsymbol{\Gamma}_{is}\boldsymbol{X}_{iqt}$$

Where $s = D_{iqt}$. The eigenvalues of $\mathbf{\Phi}_{is}$ are restricted to be within unit circle.

(4). $\boldsymbol{\gamma}_{ils} \mid \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, \boldsymbol{R}_i, \boldsymbol{\Pi}_{ls}, \boldsymbol{\Psi}_{ls} \sim MVN(\boldsymbol{B}\cdot\boldsymbol{A}, \boldsymbol{B})\cdot$

I($W^a_{iqtS} \geq W^a_{iqt(S-1)} \geq \cdots \geq W^a_{iqt1}$ and $W^b_{iqt1} \geq W^b_{iqt2} \geq \cdots \geq W^b_{iqtS}$)

Where $\boldsymbol{B} = (\sum_{q=1}^{Q_i} \sum_{t=1}^{T_{iq}} \boldsymbol{X}_{iqlt} \boldsymbol{\Sigma}_s^{-1} \boldsymbol{X}_{iqlt}^{'} + \boldsymbol{\Psi}_{ls}^{-1})^{-1}$

$$\boldsymbol{A} = \sum_{q=1}^{Q_i} \sum_{t=1}^{T_{iq}} \boldsymbol{X}_{iqlt} \boldsymbol{\Sigma}_s^{-1} \boldsymbol{U}_{iqt}^{*} + \boldsymbol{\Psi}_{ls}^{-1} \boldsymbol{\Pi}_{ls} \boldsymbol{R}_i$$

$$\boldsymbol{U}_{iqt}^{*} = \boldsymbol{U}_{iqt} - \boldsymbol{\Phi}_{is} \boldsymbol{U}_{iq(t-1)} - \boldsymbol{\Gamma}_{i(-l)s} \boldsymbol{X}_{iq(-l)t}$$

(5). $\boldsymbol{\Pi}_{kls} \mid \{\boldsymbol{\gamma}_{ils}\}, \{\boldsymbol{R}_i\}, \boldsymbol{\Psi}_{ls}, \overline{\boldsymbol{\Pi}_s}, \boldsymbol{V}_{\boldsymbol{\Pi}} \sim MVN(\boldsymbol{V} \cdot \boldsymbol{F},\ \boldsymbol{V})$

Where $\boldsymbol{V} = (\sum_{i=1}^{I} R_{ik} \boldsymbol{\Psi}_{ls}^{-1} R_{ik}^{'} + \boldsymbol{V}_{\boldsymbol{\Pi}}{}^{-1})^{-1}$, $\boldsymbol{F} = \sum_{i=1}^{I} R_{ik} \boldsymbol{\Psi}_{ls}^{-1} \boldsymbol{\gamma}_{ils}^{*} + \boldsymbol{V}_{\boldsymbol{\Pi}}{}^{-1} \overline{\boldsymbol{\Pi}_s}$, and

$\boldsymbol{\gamma}_{ils}^{*} = \boldsymbol{\gamma}_{ils} - \boldsymbol{\Pi}_{l(-k)s} \boldsymbol{R}_{i(-k)}$

(6)

$$\boldsymbol{\Psi}_{ls}{}^{-1} \mid \{\boldsymbol{\gamma}_{ils}\}, \{\boldsymbol{\Pi}_{ls}\}, \{\mathbf{R}_i\}, \rho_{\varsigma}, \mathbf{V}_{\varsigma} \sim Wishart(I + \rho_{\varsigma},\ (\mathbf{V}_{\varsigma} + \sum_{i=1}^{I} (\boldsymbol{\gamma}_{ils} - \boldsymbol{\Pi}_{ls}\mathbf{R}_i)(\boldsymbol{\gamma}_{ils} - \boldsymbol{\Pi}_{ls}\mathbf{R}_i)')^{-1})$$

(7).

$$\boldsymbol{\Sigma}_s{}^{-1} \mid \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{\Phi}_{is}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iqt}\}, \rho_{\boldsymbol{\Sigma}}, \boldsymbol{V}_{\boldsymbol{\Sigma}}$$

$$\sim Wishart(\sum_{i=1}^{I} \sum_{q=1}^{Q_i} T_{iq} + \rho_{\Sigma},$$

$$(\boldsymbol{V}_{\Sigma} + \sum_{i=1}^{I} \sum_{q=1}^{Q_i} \sum_{t=1}^{T_{iq}} (\boldsymbol{U}_{iqt} - \boldsymbol{\Phi}_{is} \boldsymbol{U}_{iq(t-1)} - \boldsymbol{\Gamma}_{is} \boldsymbol{X}_{iqt})(\boldsymbol{U}_{iqt} - \boldsymbol{\Phi}_{is} \boldsymbol{U}_{iq(t-1)} - \boldsymbol{\Gamma}_{is} \boldsymbol{X}_{iqt})')^{-1})$$

(8). $\lambda_{is} \mid \{\boldsymbol{D}_{iq}\}, shape_{prior}, scale_{prior} \sim Gamma(\sum_{q=1}^{Q_i} n_{iqs} + shape_{prior}, \sum_{q=1}^{Q_i} m_{iqs} + scale_{prior})$

where n_{iqs} is the number of times $\mathbf{D}_{iq}$ was in state s in session q and m_{iqs} is the amount of time that $\mathbf{D}_{iq}$ was in state s in session q.

(9). $\boldsymbol{D}_{iq} \mid \{\boldsymbol{\Phi}_{is}\}, \{\boldsymbol{U}_{iqt}\}, \{\boldsymbol{Y}_{iqt}\}, \{\boldsymbol{\Gamma}_{is}\}, \{\boldsymbol{X}_{iqt}\}, \boldsymbol{\Sigma}_s, \boldsymbol{v}_i, \boldsymbol{P}_{ij}$ ~ Reversible Jump Algorithm: independence sampler, refinement sampler, and birth-and-death sampler.

We use the reversible jump Hasting Metropolis (HM) algorithms proposed by Liechty and Roberts (2001) to generate samples of each hidden Markov chain $\boldsymbol{D}_{iq}$. The difference between their algorithm and ours is based on the distribution of **U** in this paper versus the likelihood functions in theirs. We used three different algorithms for updating $\boldsymbol{D}_{iq}$. The first algorithm is an independence algorithm, which ignores the current realization of $\boldsymbol{D}_{iq}$ and proposes realizations by drawing from the prior density of $\boldsymbol{D}_{iq}$. This results in proposed realizations that are considerably different, in terms of the posterior density, and as a consequence this algorithm tends to result in large but infrequent moves.

The other two algorithms create proposed realizations of $\boldsymbol{D}_{iq}$ by making small modifications to the current realization of $\boldsymbol{D}_{iq}$. The second algorithm is a refinement algorithm where the proposed realization of $\boldsymbol{D}_{iq}$ is created by modifying one of the jump times of the current realization of $\boldsymbol{D}_{iq}$. The third algorithm is a birth-death algorithm where the proposed realization of $\boldsymbol{D}_{iq}$ is created by either inserting a new interval into the current realization of $\boldsymbol{D}_{iq}$ – a birth – or removing an interval from the current realization of $\boldsymbol{D}_{iq}$ – a death. The independence algorithm has obvious advantages when the posterior distribution is multi-modal or when a poor initial value of $\boldsymbol{D}_{iq}$ has been chosen, where as the refinement algorithm and the birth-death algorithm have the advantage of more efficiently exploring the modes of the posterior distribution.

In order to take advantage of the properties of these three algorithms, one of these three algorithms is randomly chosen at each iteration of the MCMC algorithm to update each hidden Markov chain. Although our model itself is different from theirs, we apply the algorithms proposed by Liechty and Roberts (2001) and refer to their description of the algorithms and the formulas for calculating the acceptance probabilities.

(10). $\boldsymbol{v}_i \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{\alpha} \sim \, Dirichlet(\alpha_1 + \sum_{q=1}^{Q_i} d_{iq1}, \ldots, \alpha_S + \sum_{q=1}^{Q_i} d_{iqS}) \text{ with } \sum_{j=1}^{S} v_{ij} = 1$

$$where\ d_{iqs} = \begin{cases} 1 & \text{if starting state is } s \text{ for user } i \text{ in session } q \\ 0 & \text{otherwise} \end{cases}$$

(11). $\boldsymbol{P}_{ij} \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{v}_i, \boldsymbol{\tau}_j \sim Dirichlet(\tau_{j1} + \sum_{q=1}^{Q_i} m_{iqj1}, \ \ldots \ , \tau_{jS} + \sum_{q=1}^{Q_i} m_{iqjS}) \text{ with } \sum_{k=1}^{S} P_{ijk} = 1$

where m_{iqjk} is the number of jumps from state j to state k for user i in session q.

The draws of $\boldsymbol{P}_{ij}$ can be sampled from Gamma distribution with $shape = \tau_{jk} + \sum_{q=1}^{Q_i} m_{iqjk}$ and scale = 1 for all k. Then normalize each draw using the sum of all of the draws.

(12). $scale_{prior} \mid \{\boldsymbol{D}_{iq}\}, \ shape_{prior}, \lambda_{is}, shape_{\lambda}, scale_{\lambda} \sim$

$$Gamma(S \cdot I \cdot shape_{prior} + shape_{\lambda}, \ \sum_{i=1}^{I}\sum_{s=1}^{S} \lambda_{is} + scale_{\lambda})$$

(13). $shape_{prior} \mid \{\boldsymbol{D}_{iq}\}, \lambda_{\text{is}}\ , scale_{prior}, shape_{\lambda}, scale_{\lambda}$

$$\propto \left\{ \prod_{i=1}^{I} \prod_{s=1}^{S} \frac{scale_{prior}{}^{shape_{prior}}}{\Gamma(shape_{prior})} \cdot \lambda_{\text{is}}{}^{shape_{prior}-1} \cdot \mathrm{e}^{-scale_{prior}*\lambda_{\text{is}}} \right\} \cdot$$

$$\frac{scale_{\lambda}{}^{shape_{\lambda}}}{\Gamma(shape_{\lambda})} \cdot shape_{prior}{}^{shape_{\lambda}-1} \cdot e^{-shape_{prior}*scale_{\lambda}}$$

$shape_{prior}$ can be generated using standard Metropolis-Hasting algorithm.

(14). $\alpha_s \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{v}_i, shape_{\alpha}, scale_{\alpha}$

$$\propto \left\{ \prod_{i=1}^{I} \frac{\Gamma(\sum_{\text{k}=1}^{\text{S}} \alpha_{\text{k}})}{\Gamma(\alpha_{\text{s}}) \cdot \Gamma(\sum_{\text{k}=1}^{\text{S}} \alpha_{\text{k}} - \alpha_{\text{s}})} \cdot v_{\text{is}}{}^{\alpha_{\text{s}}-1} \cdot (1 - v_{\text{is}})^{\sum_{\text{k}=1}^{\text{S}} \alpha_{\text{k}} - \alpha_{\text{s}} - 1} \right\} \cdot$$

$$\frac{scale_{\alpha}{}^{shape_{\alpha}}}{\Gamma(shape_{\alpha})} \cdot \alpha_{\text{s}}{}^{shape_{\alpha}-1} \cdot e^{-\alpha_{\text{s}}*scale_{\alpha}}$$

α_s can be generated using standard Metropolis-Hasting algorithm.

(15). $\tau_{sj} \mid \{\boldsymbol{D}_{iq}\}, \boldsymbol{P}_{ij}, shape_{\tau}, scale_{\tau}$

$$\propto \left\{ \prod_{i=1}^{I} \frac{\Gamma(\sum_{\text{k}=1}^{\text{S}} \tau_{kj})}{\Gamma(\tau_{sj}) \cdot \Gamma(\sum_{\text{k}=1}^{\text{S}} \tau_{kj} - \tau_{sj})} \cdot p_{\text{ijs}}{}^{\tau_{sj}-1} \cdot (1 - p_{\text{ijs}})^{\sum_{\text{k}=1}^{\text{S}} \tau_{kj} - \tau_{sj} - 1} \right\} \cdot$$

$$\frac{scale_{\tau}{}^{shape_{\tau}}}{\Gamma(shape_{\tau})} \cdot \tau_{sj}{}^{shape_{\tau}-1} \cdot e^{-\tau_{sj}*scale_{\tau}}$$

τ_{sj} can be generated using standard Metropolis-Hasting algorithm.

Appendix C: Categorization of Web Pages

We assign pages to one of seven categories as summarized in Table A. The assignment of URLs to a category was made by scanning for specific keywords that occur in the URL string. The keywords used to identify each category are given in Table B. For example, the string "/index.asp" occurs in the URL http://www.bn.com/index.asp, hence this URL would be assigned to the home category. The keywords and categories are given in the order of precedence. These categories were motivated by the navigational structure of the B&N site and research in human computer interaction (Redish 2002). To illustrate our data we report the data collected by CMM for a user session in Table C.

Abbr.	Category	Description of Page Content
H	Home	Home page, common starting page for B&N visitors
A	Account	User sign in, address changes, and review order status
C	Category	List of items (category listings) or search results
P	Product	Detailed product information, item description, price information, availability, and reviews
I	Information	Shipping, order status and popup advertisements
S	Shopping Cart	Reviewing the cart, delete items, entering purchase information
O	Order	Confirmation page that denotes order has taken place
E	Enter/Exit	Non B&N pages used to denote the beginning or end of a session

Table A. Categorization scheme for B&N web pages for 9,180 unique pages requests.

Category Abbreviation	Category Name	Keywords
A	Account	account/, youraccount.asp, signin.asp
C	Category	/subjects/, /bookstore.asp, /oopbooks/, /textbooks/, /bargain/, /ebooks/, music/, music/index.asp, video/, video/index.asp, posters/, posters/index.asp, enews/main/, /boutiques/, /holiday.asp, /genre.asp, /topshelf/, /bookshelf/, /style.asp, /newreleases/, /home.asp, /collection/, /bestsellers/, /comingsoon/, /recommended/, /firstchapters\, /booksearch/results.asp, /education/, /onthecharts/, /recommend/, /classical/, /budget/, /nr_home.asp, /topsell.asp, enews/category/, enews/browse/, enews/gifts/, enews/newspapers/
P	Product	booksearch/isbninquiry.asp, enews/magazine/, /search/product.asp
S	Shopping Cart	shop/cart.asp, shop/shipping.asp, shop/giftwrapping.asp, shop/payment.asp, shop/billing.asp
O	Order	shop/confirm.asp
H	Home	bn.com/index.asp, barnesandnoble.com/index.asp
I	Information	All other URLs not classified in other categories
E	Enter/Exit	URLs not containing "bn.com" or "barnesandnoble.com"

Table B. Keywords used to assign URLs to appropriate category.

	Time	URL	Category	Abbr.
1	8:36:11pm	/promo/coupon/popups/fs_usa_popup.asp?userid=xxx	Information	I
2	8:36:29pm	/booksearch/results.asp?wrd=70%2d215&userid=xxx	Category	C
3	8:36:48pm	/booksearch/results.asp?userid=xxx&mscssid=yyy&wrd=70%2d215&opr=a&sort=p	Category	C
4	8:37:14pm	/booksearch/isbninquiry.asp?userid=xxx&mscssid=yyy&isbn=0072134445	Product	P
5	8:38:10pm	/booksearch/results.asp?userid=xxx&mscssid=yyy&wrd=70%2d215&opr=a&sort=p	Category	C
6	8:44:32pm	/textbooks/booksearch/isbninquiry.asp?userid=xxx&mscssid=yyy&isbn=0619034971	Product	P
7	8:55:12pm	/promo/coupon/popups/fs_usa_popup.asp?userid=xxx	Information	I
8	8:55:24pm	/booksearch/results.asp?wrd=70%2d215&userid=xxx	Category	C
9	8:55:36pm	/booksearch/results.asp?userid=xxx&mscssid=yyy&wrd=70%2d215&opr=a&sort=p	Category	C
10	8:56:37pm	/shop/signin.asp?userid=xxx&mscssid=yyy	Account	A
11	8:58:16pm	/booksearch/results.asp?userid=xxx&mscssid=yyy&wrd=70%2d215&opr=a&sort=p	Category	C
12	8:58:40pm	/booksearch/isbninquiry.asp?userid=xxx&mscssid=yyy&isbn=0072224983	Product	P
13	8:59:21pm	/shop/cart.asp?userid=xxx&mscssid=yyy	Shop Cart	S
14	9:01:26pm	Exit	Exit	E

Table C. Listing of raw clickstream dataset associated with a selected session on April 28, 2002 for one user. (All URLs are prefixed by http://www.barnesandnoble.com, and the userid and mscssid is listed as xxx and yyy, respectively, to protect the privacy of this user.)

Appendix D: Parameter Estimates

Impact of Marketing Mix, Hypertext Links, and Behavioral Variables

The posterior means and standard deviations of the hyper-distribution parameter estimates for the marketing mix and hypertext link covariates of our two-state dynamic multinomial probit model are given in Table 1 and 2, respectively (in our notation this is $\mathrm{E}[\boldsymbol{\Gamma}]$ in equation 5, where the demographics are set to the mean values).

The intercepts provide the basic attractiveness of a category holding everything else constant. Notice that users in a deliberation-oriented state tend to have more positive intercepts than those in the browsing-oriented state, which indicates the deliberation-oriented visitors will have longer sessions over browsing-oriented visitors. The most likely category to be viewed by browsing-oriented users is the home page category, while visitors in a deliberation-oriented state are more likely to visit account pages. Perhaps users in a browsing-oriented state are less likely to delve deeply into a web site compared to deliberation-oriented users.

		Inter-cept	Price Present	Promo-tion	Ads	Home Links	Acct Links	Product Links
Browsing-oriented State	Home	.29	-.10	.05	-.15	-.10	-.06	-.01
		(.08)	(.03)	(.01)	(.03)	(.02)	(.01)	(.01)
	Account	-.09	-.17	-.04	.20	-.04	.10	.01
		(.02)	(.03)	(.02)	(.03)	(.01)	(.01)	(.01)
	Category	-.01	.10	.01	-.17	-.01	-.01	.03
		(.02)	(.02)	(.01)	(.03)	(.01)	(.01)	(.01)
	Product	-.07	.12	-.08	.08	.02	.03	.01
		(.02)	(.03)	(.01)	(.02)	(.01)	(.01)	(.01)
	Information	-.06	-.02	.04	.05	.03	.03	.01
		(.03)	(.02)	(.02)	(.03)	(.01)	(.01)	(.01)
	Shopping Cart	-.11	-.07	.05	-.04	-.04	.01	.01
		(.02)	(.02)	(.01)	(.02)	(.01)	(.01)	(.01)
	Order	-.54	-.04	.05	-.06	-.02	.04	.01
		(.08)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Deliberation-oriented State	Home	-.35	-.02	.03	-.07	.03	-.01	-.01
		(.14)	(.01)	(.01)	(.03)	(.01)	(.01)	(.01)
	Account	.61	-.03	-.04	.01	-.01	.02	-.02
		(.13)	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)
	Category	.07	.03	-.01	.05	.02	.01	.06
		(.08)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
	Product	-.15	.26	.03	-.09	-.01	.03	.07
		(.06)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
	Information	.01	-.01	.02	.13	.06	.03	.02
		(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
	Shopping Cart	.27	.04	-.14	-.03	.01	.06	.04
		(.03)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
	Order	.06	.08	-.03	-.07	.01	-.03	.02
		(.02)	(.02)	(.01)	(.01)	(.01)	(.01)	(.01)

Table 1. Posterior mean and standard deviation of parameter estimates for marketing mix and hypertext links.

State	Category	Purchase in Last Session	Time Duration	Visit Depth	Weekend	Other Site Visits	Other Bookstores
Browsing-oriented State	Home	.29	.04	.09	.29	.19	.04
		(.05)	(.02)	(.02)	(.05)	(.03)	(.05)
	Account	-.19	-.05	.04	.15	-.12	.16
		(.05)	(.02)	(.01)	(.04)	(.03)	(.07)
	Category	.33	-.08	-.07	.11	-.13	-.14
		(.04)	(.02)	(.01)	(.03)	(.03)	(.04)
	Product	.10	.05	-.02	.01	.05	.10
		(.05)	(.02)	(.01)	(.02)	(.03)	(.04)
	Information	.21	-.11	-.02	.26	.01	-.37
		(.05)	(.01)	(.02)	(.04)	(.02)	(.05)
	Shopping Cart	-.13	-.09	.03	.15	-.11	-.36
		(.06)	(.01)	(.01)	(.03)	(.02)	(.06)
	Order	.33	.03	.01	.02	.01	-.02
		(.05)	(.02)	(.01)	(.03)	(.01)	(.01)
Deliberation-oriented State	Home	.05	-.04	.02	.02	.01	.09
		(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
	Account	-.01	-.01	.02	-.01	.04	-.20
		(.01)	(.01)	(.01)	(.01)	(.01)	(.04)
	Category	-.03	.01	-.31	.01	-.01	-.02
		(.01)	(.01)	(.03)	(.01)	(.01)	(.01)
	Product	.06	-.01	-.02	.01	-.05	-.04
		(.02)	(.01)	(.01)	(.01)	(.01)	(.01)
	Information	.02	.02	-.04	.01	-.01	.02
		(.01)	(.01)	(.01)	(.01)	(.02)	(.01)
	Shopping Cart	.01	.02	.04	.05	.02	-.05
		(.01)	(.01)	(.01)	(.01)	(.02)	(.01)
	Order	.03	-.01	.05	-.05	-.01	.04
		(.01)	(.01)	(.01)	(.01)	(.01)	(.01)

Table 2. Posterior mean and standard deviation of parameter estimates for web browsing context variables.

Estimates of the VAR and Markov Model of the Mixture Process

The estimates of the hyper-distribution for the VAR(1) process (denoted as $\overline{\Phi}_s$ in equation 7) are given in table 3. We find that although these coefficients are relatively small, there are significant time trends in all categories. Positive diagonal elements among account, category, and product pages indicate that previously viewing a page in this category increases the probability of viewing another page in this category. While negative parameters indicate a higher chance of not viewing the category again, such as the home, shopping cart, and order pages. The diagonal elements in the deliberation-oriented state range from -.09 to .23, while in the browsing-oriented state they range from -.08 to .08, indicating slightly more persistence in the deliberation-oriented-state.

Positive off-diagonal terms indicate switching behavior from one category to another, while negative off-diagonal terms indicate a decreased likelihood of switching from one category to the next. For example, users who have previously viewed a product page become more likely to view a home or account page when the latent utility of these corresponding categories increase.

State	Categories	Home	Account	Category	Product	Inform	ShopCart	Order
Browsing-oriented State	Home	-.08	.27	-.03	.18	.04	-.07	-.79
		(.020)	(.002)	(.003)	(.003)	(.014)	(.012)	(.016)
	Account	.02	.05	-.08	.22	-.12	.39	.35
		(.015)	(.002)	(.004)	(.005)	(.014)	(.018)	(.010)
	Category	.01	-.01	.08	.06	.23	-.06	.47
		(.012)	(.001)	(.001)	(.002)	(.009)	(.011)	(.007)
	Product	-.03	-.03	.03	.06	-.09	.44	.17
		(.008)	(.001)	(.001)	(.003)	(.003)	(.032)	(.004)
	Information	.01	-.13	-.02	-.03	.03	.11	.34
		(.014)	(.001)	(.001)	(.001)	(.001)	(.014)	(.003)
	Shopping Cart	.02	.09	-.01	-.02	-.02	-.01	.50
		(.012)	(.001)	(.001)	(.001)	(.002)	(.013)	(.003)
	Order	-.01	.13	-.05	-.02	-.01	.03	-.01
		(.008)	(.002)	(.001)	(.001)	(.001)	(.001)	(.002)
Deliberation-oriented State	Home	-.09	.05	-.47	.39	.31	-.68	.08
		(.004)	(.004)	(.041)	(.002)	(.013)	(.016)	(.007)
	Account	-.03	.04	-.06	.43	.24	-.14	.20
		(.008)	(.006)	(.004)	(.004)	(.017)	(.019)	(.009)
	Category	-.03	-.03	.23	-.06	-.15	.82	-.07
		(.004)	(.003)	(.006)	(.002)	(.006)	(.008)	(.032)
	Product	-.01	-.01	.04	.08	-.13	.40	.09
		(.003)	(.005)	(.005)	(.007)	(.007)	(.027)	(.003)
	Information	-.01	-.03	-.01	-.01	-.02	-.44	-.49
		(.003)	(.003)	(.005)	(.003)	(.001)	(.011)	(.006)
	Shopping Cart	.01	.02	-.06	.01	-.01	-.01	.16
		(.002)	(.003)	(.004)	(.001)	(.001)	(.001)	(.006)
	Order	.02	.02	.01	-.03	-.01	.03	-.02
		(.003)	(.002)	(.004)	(.001)	(.001)	(.001)	(.002)

Table 3. Posterior mean and standard deviation of the VAR parameter hyper-distribution.

The estimates for the two-state hidden Markov chain are given in Table 4. Notice that a user has a high probability of starting in a browsing-oriented state (64%). On average a user will stay in this browsing-oriented state for about three viewings (i.e., the inverse of waiting time is

.32). In contrast, a user in the deliberation-oriented state tends to persist in this state longer or about four viewings (i.e., the inverse of waiting time is .26). The transition probability matrix is trivial for the two-state model, since there are only two states and the switching behavior is captured by the waiting time in each state.

	Browsing oriented state	Deliberation oriented state
λ (Inverse of Waiting Time)	.32	.26
	(.01)	(.01)
ν (Starting Probabilities)	.64	.36
	(.01)	(.01)
P (Transition Probabilities)	0	1
	1	0

Table 4. Posterior mean and standard deviation of parameters associated with state transitions.

Unexplained variance in Web Page Selection

The estimates of the error covariance matrix ($\boldsymbol{\Sigma}_s$; upper triangular elements in the table) and correlations (lower triangular) are given in Table 5. Notice that the variance for the home page category is set to 1 for identification purposes. Consistent with our prior expectations most of the off-diagonal covariance estimates in both the browsing-oriented state and deliberation-oriented state are significantly different than zero. This indicates that the independence assumption of the independent probit model is incorrect. Also we find that the variance estimates in the browsing-oriented state are slightly larger than those of the purchase-oriented state, which indicates that these users are less persistent in category choices.

Substantively, we find that users in browsing-oriented states who unexpectedly choose account pages are more likely to choose category pages after controlling for covariates and lagged effects. In contrast, users in a deliberation-oriented state appear to be less likely to move from an account to a category page. Notice that deliberation-oriented consumers are less likely to move from an information page to a home, account, category, or shopping cart page, while browsing-oriented consumers are somewhat more likely.

State	Categories	Home	Account	Category	Product	Inform	ShopCart	Order
Browsing-oriented State	Home	1	.074	.047	.073	.030	.017	.014
	Account	.140 (.003)	3.535 (.166)	.229	.053	.082	.001	.007
	Category	.042 (.002)	.380 (.014)	0.776 (.038)	.129	.028	.0480	-.011
	Product	.060 (.002)	.082 (.004)	.094 (.022)	0.685 (.034)	.019	.054	.017
	Information	.023 (.001)	.119 (.004)	.019 (.006)	.012 (.010)	0.586 (.020)	.046	.001
	Shopping Cart	.012 (.008)	.002 (.005)	.290 (.008)	.030 (.006)	.024 (.005)	0.468 (.016)	.001
	Order	.090 (.004)	.180 (.008)	-.060 (.003)	.090 (.002)	.011 (.003)	.010 (.002)	0.414 (.013)
Deliberation-oriented State	Home	1	-.213	.153	.028	-.186	.109	-.016
	Account	-.333 (.005)	2.440 (.322)	-.109	.050	-.152	.161	.164
	Category	.127 (.002)	-.142 (.010)	0.688 (.089)	-.109	-.288	.105	.054
	Product	.021 (.002)	.058 (.012)	-.068 (.006)	0.556 (.069)	.239	.041	.009
	Information	-.155 (.003)	-.198 (.008)	-.199 (.008)	.149 (.068)	0.697 (.094)	-.200	.056
	Shopping Cart	.065 (.002)	.151 (.005)	.052 (.004)	.018 (.034)	-.100 (.040)	0.361 (.031)	-.040
	Order	-.008 (.008)	.121 (.002)	.021 (.008)	.003 (.009)	.022 (.008)	-.010 (.006)	0.223 (.021)

Table 5. Estimated error covariance matrices for the two-state model.

Impact of Demographics on Web Browsing

The demographic variables are meant to help explain variation in browsing behavior across users. The posterior mean and standard deviation for the hyper-parameters associated with the impact of the demographic variables on the web user's response to price presence are given in Table 6 (we denote these parameters as $\mathbf{\Pi}_{ls}$ in equation 6), and the associated covariance matrix ($\mathbf{\Psi}_s$) is given in Table 7. For brevity we report the demographic relationships for only whether price information is present on the page (price presence), but have estimated all demographic responses.

State	Categories	Inter-cept	Age	Age^2	Male	White	Child	Married	College	Med. Income	High Income
Browsing-oriented State	Home	-.17	.04	-.02	.42	-.15	.25	.01	-.03	.24	.19
		(.03)	(.02)	(.02)	(.02)	(.02)	(.03)	(.01)	(.01)	(.02)	(.02)
	Account	-.58	.21	-.01	-.02	-.05	.33	-.20	-.30	-.44	-.20
		(.02)	(.01)	(.01)	(.02)	(.01)	(.03)	(.01)	(.03)	(.03)	(.02)
	Category	1.32	.14	.01	.16	.12	-.10	.10	-.38	.24	.59
		(.03)	(.01)	(.01)	(.01)	(.02)	(.01)	(.02)	(.03)	(.02)	(.03)
	Product	.89	-.30	.05	.09	-.04	-.26	.18	.23	-.51	.13
		(.03)	(.01)	(.01)	(.02)	(.02)	(.01)	(.02)	(.01)	(.02)	(.02)
	Information	-1.21	.24	-.04	.37	-.40	.47	.14	.65	-.18	-.20
		(.02)	(.01)	(.01)	(.02)	(.03)	(.02)	(.03)	(.02)	(.02)	(.02)
	Shopping Cart	-.48	.04	.01	.06	-.17	.26	.24	.22	.21	.03
		(.02)	(.01)	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)	(.02)	(.02)
	Order	-.32	.16	-.02	.18	-.32	.07	.06	.05	.02	.07
		(.02)	(.01)	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)	(.02)	(.02)
Deliberation-oriented State	Home	-.04	-.01	-.01	-.01	-.01	.03	-.16	.06	.03	.14
		(.02)	(.02)	(.01)	(.02)	(.02)	(.02)	(.03)	(.02)	(.03)	(.03)
	Account	-.02	-.01	-.01	.13	.02	-.01	-.02	.04	-.01	.03
		(.02)	(.01)	(.01)	(.02)	(.02)	(.01)	(.02)	(.03)	(.02)	(.03)
	Category	.02	.01	.01	-.01	-.09	.02	-.03	.05	.01	-.09
		(.01)	(.03)	(.01)	(.02)	(.02)	(.01)	(.02)	(.03)	(.02)	(.03)
	Product	.02	.01	-.01	-.08	.01	-.03	-.12	-.11	-.15	.12
		(.01)	(.01)	(.01)	(.02)	(.02)	(.01)	(.03)	(.03)	(.02)	(.02)
	Information	-.01	-.01	.02	-.21	-.15	.04	-.03	.04	.08	-.23
		(.01)	(.02)	(.01)	(.03)	(.02)	(.01)	(.02)	(.03)	(.02)	(.03)
	Shopping Cart	.03	.01	.01	.01	-.07	.05	-.01	-.12	.10	.13
		(.01)	(.02)	(.01)	(.02)	(.02)	(.02)	(.02)	(.03)	(.02)	(.03)
	Order	.03	.01	.01	-.10	.04	.03	-.19	-.09	.04	-.01
		(.01)	(.02)	(.01)	(.03)	(.02)	(.02)	(.03)	(.03)	(.02)	(.02)

Table 6. Posterior mean and standard deviation of effects of demographics on browsing.

State	Categories	Home	Account	Category	Product	Inform	Shop Cart	Order
Browsing-oriented State	Home	.00019 (.00003)	.47	.55	.38	.55	.48	.89
	Account	.00008 (.00001)	.00015 (.00004)	.41	.57	.48	.16	.93
	Category	.00009 (.00003)	.00006 (.00003)	.00014 (.00001)	.67	.14	.81	.71
	Product	.00006 (.00003)	.00008 (.00003)	.00009 (.00002)	.00013 (.00003)	.67	.59	.54
	Information	.00009 (.00002)	.00007 (.00002)	.00002 (.00002)	.00009 (.00002)	.00014 (.00001)	.32	.78
	Shopping Cart	.00007 (.00002)	.00002 (.00001)	.00010 (.00001)	.00007 (.00002)	.00004 (.00002)	.00011 (.00002)	.44
	Order	.00016 (.00002)	.00015 (.00003)	.00011 (.00003)	.00008 (.00002)	.00012 (.00002)	.00006 (.00001)	.00017 (.00003)
Deliberation-oriented State	Home	.00015 (.00004)	.37	-.94	.74	.43	.50	.67
	Account	.00005 (.00003)	.00012 (.00001)	-.28	.50	.55	.42	-.41
	Category	-.00015 (.00002)	-.00004 (.00001)	.00017 (.00001)	-.56	-.37	-.65	-.69
	Product	.00015 (.00004)	.00009 (.00003)	-.00012 (.00002)	.00027 (.00004)	.71	.65	.54
	Information	.00013 (.00001)	.00015 (.00002)	-.00012 (.00001)	.00029 (.00004)	.00062 (.00004)	.43	.06
	Shopping Cart	.00008 (.00001)	.00006 (.00001)	-.00011 (.00005)	.00014 (.00011)	.00014 (.00001)	.00017 (.00001)	.74
	Order	.00011 (.00001)	-.00006 (.00001)	-.00012 (.00001)	.00012 (.00001)	.00002 (.00001)	.00013 (.00001)	.00018 (.00001)

Table 7. Estimated covariance matrix of the hyper-distribution for the two-state model. The lower triangular portion of the matrix contains the estimates of the covariance matrix, and the upper triangular portion has the correlation estimates.

Appendix E: Multi-Step Forecasting Performance

The fit and hit-rate provided in Tables 7 and 8 of the paper measure one-step ahead forecasting performance. However, we are not simply interested in forecasting a single-step ahead, but we are potentially interested in predicting the entire path that a user may take. In Table 8 we report the accuracy of various models in predicting the ten most frequent paths with a specific two-category sequence. For example, in the estimation dataset, the two-category sequence "CC" occurs 2,066 times, which is more than any other two sequences. Our best model correctly predicts this sequence 73.63% of the time when it actually occurs. The task of predicting the sequence "CC" requires predicting the probability that both the one- and two-step ahead predictions are "C". This is a harder problem than a one-step ahead forecast, since the error in the first-step will be compounded in the forecast of the second step. Notice that our best model, the dynamic two-state Probit model, performs significantly better than all other models.

Sample	Model	Paths With Two Following Two Category Sequence									
		CC	II	EI	PP	CP	IC	AA	IE	PC	SS
Estimation	Actual	2066	1352	776	663	547	490	460	455	414	361
	Zero-Order Markov (1 State)	12.15	4.81	2.58	3.17	5.67	8.16	0.87	3.30	5.07	0
		(0.33)	(0.22)	(0.16)	(0.18)	(0.24)	(0.28)	(0.09)	(0.18)	(0.22)	(0.02)
	First-Order Markov (1 State)	31.50	15.63	3.76	10.76	8.39	6.08	43.04	5.93	7.49	11.91
		(0.47)	(0.37)	(0.19)	(0.32)	(0.28)	(0.24)	(0.51)	(0.24)	(0.27)	(0.33)
	Multinomial Probit	15.61	8.73	2.66	9.05	6.19	7.43	13.26	3.32	5.42	4.71
		(0.37)	(0.29)	(0.16)	(0.29)	(0.25)	(0.26)	(0.35)	(0.18)	(0.23)	(0.22)
	Latent Class (2 States)	17.13	11.17	2.94	11.28	7.18	7.41	16.45	3.87	5.78	5.44
		(0.38)	(0.32)	(0.17)	(0.32)	(0.26)	(0.27)	(0.38)	(0.20)	(0.24)	(0.23)
	VAR with Intercept	53.92	42.07	30.03	21.66	32.38	42.04	46.30	20.40	33.14	31.94
		(0.51)	(0.50)	(0.47)	(0.42)	(0.48)	(0.50)	(0.51)	(0.41)	(0.48)	(0.47)
	Dynamic Probit (1 State)	63.73	62.22	40.98	42.87	54.57	53.67	55.22	40.03	43.14	42.77
		(0.49)	(0.50)	(0.50)	(0.51)	(0.51)	(0.51)	(0.51)	(0.50)	(0.50)	(0.51)
	Dynamic Probit (2 States)	73.63	72.22	51.33	53.77	55.85	61.63	63.91	50.77	54.59	53.05
		(0.45)	(0.46)	(0.51)	(0.51)	(0.51)	(0.49)	(0.49)	(0.51)	(0.51)	(0.51)
Holdout	Actual	889	563	213	353	286	148	221	139	249	217
	Zero-Order Markov (1 State)	9.24	4.62	2.82	2.83	5.59	8.78	0	1.44	5.22	0.46
		(0.41)	(0.30)	(0.24)	(0.24)	(0.33)	(0.40)	(0.06)	(0.17)	(0.32)	(0.09)
	First-Order Markov (1 State)	29.38	12.28	3.09	12.97	8.23	3.47	39.82	2.16	7.23	13.82
		(0.65)	(0.47)	(0.25)	(0.49)	(0.39)	(0.26)	(0.70)	(0.21)	(0.37)	(0.49)
	Multinomial Probit	13.07	5.20	2.88	6.23	5.75	5.35	11.81	1.46	5.20	3.82
		(0.48)	(0.32)	(0.24)	(0.34)	(0.33)	(0.32)	(0.46)	(0.17)	(0.32)	(0.27)
	Latent Class (2 States)	15.43	15.10	2.90	14.79	7.05	4.68	15.65	1.87	6.09	4.66
		(0.51)	(0.51)	(0.24)	(0.50)	(0.36)	(0.30)	(0.52)	(0.19)	(0.34)	(0.30)
	VAR with Intercept	44.39	32.84	28.07	24.25	24.19	32.03	36.79	18.34	33.21	25.88
		(0.71)	(0.67)	(0.61)	(0.61)	(0.61)	(0.67)	(0.69)	(0.55)	(0.67)	(0.62)
	Dynamic Probit (1 State)	52.81	53.91	35.10	33.40	50.70	43.38	44.52	36.93	40.80	40.15
		(0.71)	(0.71)	(0.68)	(0.67)	(0.71)	(0.70)	(0.71)	(0.69)	(0.70)	(0.70)
	Dynamic Probit (2 States)	64.16	63.55	44.90	51.70	52.80	52.70	54.52	45.50	53.61	46.23
		(0.68)	(0.69)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)

Table 8. Hit rate (%) for 10 most frequent paths with a specific two-category sequence. The entries for the "Actual" row are frequencies while the other rows report the hit-rate of each model predicting this path. The standard errors of the hit rates are provided in parentheses below the estimate.

Appendix F: Predicting Purchase Conversion

For comparison we report the probabilities for various alternative model formulations in Table 9.

Sample	Model	Forecast Origin/Number of viewings during session					
		1	2	3	4	5	6
Estimation	First-Order Markov (1 State)	7.2% (0.26)	8.2% (0.28)	10.1% (0.31)	12.7% (0.34)	15.3% (0.36)	22.4% (0.43)
	Latent Class (2 States)	7.4% (0.27)	7.8% (0.27)	9.5% (0.30)	11.3% (0.32)	12.8% (0.34)	14.5% (0.36)
	Intercept + VAR	10.4% (0.31)	11.6% (0.33)	14.9% (0.36)	17.0% (0.38)	21.4% (0.42)	26.0% (0.45)
	Dynamic Multinomial Probit (1 State, Page-Level)	12.4% (0.34)	14.0% (0.35)	18.7% (0.39)	25.1% (0.44)	29.0% (0.46)	35.8% (0.49)
	Dynamic Multinomial Probit (2 States, Page-Level)	13.3% (0.48)	16.3% (0.52)	23.4% (0.60)	30.9% (0.65)	34.4% (0.67)	41.5% (0.69)
Holdout	First-Order Markov (1 State)	6.5% (0.35)	7.5% (0.37)	9.6% (0.42)	12.6% (0.47)	13.7% (0.49)	16.6% (0.53)
	Latent Class (2 States)	7.2% (0.82)	7.3% (0.82)	8.6% (0.89)	9.5% (0.93)	9.9% (0.94)	11.2% (0.99)
	Intercept + VAR	8.4% (0.88)	9.3% (0.92)	11.8% (1.02)	13.5% (1.08)	15.3% (1.13)	17.7% (1.21)
	Dynamic Multinomial Probit (1 State, Page-Level)	9.3% (0.92)	11.4% (1.01)	15.2% (1.13)	16.4% (1.17)	17.8% (1.21)	19.0% (1.24)
	Dynamic Multinomial Probit (2 States, Page-Level)	10.4% (0.97)	12.8% (1.06)	15.2% (1.14)	18.0% (1.21)	19.1% (1.24)	21.2% (1.29)

Table 9. Predicted purchase conversion probabilities (and standard errors in parentheses) of users who purchase given initial paths for various models. The estimation sample has 83 sessions and the holdout has 31.