Web Appendices:

From Clicks to Returns: Website Browsing and Product Returns

Marat Ibragimov
Assistant Professor of Marketing
Emory University Goizueta Business School
marat.ibragimov@emory.edu

Siham El Kihal
Assistant Professor of Marketing
Frankfurt School of Finance & Management
s.elkihal@fs.de

John R. Hauser
Kirin Professor of Marketing
MIT Sloan School of Management
hauser@mit.edu

Web Appendix A: Data Preprocessing

Overall data cleaning. Search data is typically noisy; therefore, we preprocessed the data to obtain better estimates of the model parameters. We took the following steps in the preprocessing:

1. Removed non-fashion products (e.g. linen, towels) and kid's apparel. These products constitute a small proportion of the data and are not the retailer's focus (95% of purchases are adult fashion products).
2. Removed browsing sessions without product listing views (each product listing contains up to 96 products presented to the customer, see Figure 6 for the example). This could happen if the customer comes to the website from a third-party website and lands directly on the product page. These sessions do not represent the true customer search process at the retailer's website and we are not able to recover the set of products from which the customer was choosing.
3. Removed browsing sessions that have more than 50 pages viewed, products clicked or products purchased.
4. Removed browsing sessions where customers were viewing product listings of size greater than 48. Our retailer allows the customer to view 48 or 96 products in one listing, most customers (and the default option is to view 48 products).
5. Removed browsing sessions that have not clicked products after a page view and sessions that have clicked products before a page view. This implies we kept only sessions with the clean search process: the customer views the product page and selects a product to click on it. The alternative could happen if the customer found a product through an

alternative means (from a third-party website) and in this case, it is impossible to infer the set of products from which he or she was choosing.

Selecting single-item orders. In the paper, we consider orders where the customer purchased at most one product. However, there are two additional steps used to obtain the representative data sample:

1. Sessions without a purchase. After we selected transactions with only one item purchased, we randomly subsampled sessions without a purchase to preserve the relative purchase rate.
2. Orders with one product but multiple sizes or several identical units. In this case, we split the order into several orders with the same search session, and only one unit was purchased, however, these orders could have different return outcomes. This approach allows us to keep more data and thus improve the estimation quality. Using alternative approaches does not lead to substantial changes.

The preprocessing procedures do not change the main message of the paper and are aimed at obtaining a representative data sample, which would balance the quality and quantity of data. In practice, the retailer may implement different preprocessing procedures which could change the parameter values, but qualitative findings would remain similar.

Web Appendix B: Deep-Learning Embeddings

In the paper, we mentioned that during the estimation we used deep-learning product embeddings to address the issue of high dimensionality of the data. Our procedure for extracting the product embeddings could be summarized in the following steps:

1. Creating product base features:
    a. Combine product quantitative characteristics (category dummy, price, brand).
    b. Use ResNet model to generate product image embeddings (2048-dimensional vectors) and PCA transformation to extract 64 components.
    c. Concatenate (a) and (b).
2. Computing aggregate product-level outcomes:
    a. Click rate $\mathrm{cr_j}$ – the ratio of clicks to views.
    b. Purchase rate $\mathrm{pr_j}$ – the ratio of sales to clicks.
    c. Return rate $\mathrm{rr_j}$ – the ratio of returns to sales.
3. Training the neural network to produce 7-dimensional product embeddings $\mathrm{e_j}$ such that $\mathbf{e}_j'\boldsymbol{\beta}^{\mathrm{cr}}$, $\mathbf{e}_j'\boldsymbol{\beta}^{\mathrm{pr}}$, $\mathbf{e}_j'\boldsymbol{\beta}^{\mathrm{rr}}$ minimize the prediction error of log-ratio of $\mathrm{cr_j}$, $\mathrm{pr_j}$, $\mathrm{rr_j}$ respectively.

These embeddings represent the product in terms of their three key characteristics, click rate, purchase rate, and return rate. There exist other ways to construct product embeddings, however, exploring these options goes beyond the scope of the paper. We leave this exercise to the retailer who may have different available data. We only note that this approach is sufficient to support the main message of the paper.

Web Appendix C: Derivation of Expected Purchase Utility

Without loss of generality, we drop all indices and subscripts in this section to preserve readability. In the section discussing the model, we wanted to find the expected utility of purchasing a product from the website. In this case, the customer knows the variables μ, ϵ and ξ and computes the expected utility in Equation 1 over ψ:

$$\omega = \mathbb{E}_{\psi}[(\mu + \xi + \epsilon + \psi + R) \cdot (\mu + \xi + \epsilon + \psi + R \geq 0) - R|\mu, \xi, \epsilon]$$
$$\mathbb{E}_{\psi}[\zeta \cdot (\zeta \geq 0)|\mu, \xi, \epsilon] - R = \sigma_{\psi} \cdot T\left(\frac{\mu + \xi + \epsilon + R}{\sigma_{\psi}}\right) - R \quad \text{(W1)}$$

where $\zeta|\mu, \xi, \epsilon \sim \mathcal{N}(\mu + \xi + \epsilon + R; \sigma_{\psi})$ and formula for the expectation of truncated normal distribution was used.

Web Appendix D: Derivation of Reservation Utilities for Model with Product Returns

In the original paper, Weitzman (1979) demonstrated that the reservation utility z for a product could be found from Equation W2 where we drop the individual i and produc j indices for compactness:

$$c = \int_{z}^{\infty} (u - z) dF(u) \quad \text{(W2)}$$

In the section discussing the model, we demonstrated that the return option changes the distribution of the reward and thus, in this case, we need to find the distribution of the expected purchase utility from Equation W1. Notice that before making a click, the customer observes only μ and ξ, therefore the randomness in Equation W1 comes from the post-click preferences shock ϵ:

$$F(u) = \mathbb{P}[\omega(\epsilon) \leq u|\mu, \xi] = \mathbb{P}\left[\sigma_{\psi} \cdot T\left(\frac{\mu + \xi + \epsilon + R}{\sigma_{\psi}}\right) - R \leq u|\mu, \xi\right]$$
$$= \mathbb{P}\left[\tilde{\mu} + \epsilon + R \leq \sigma_{\psi} T^{-1}\left(\frac{R + u}{\sigma_{\psi}}\right)|\mu, \xi\right] = \Phi\left[\sigma_{\psi} T^{-1}\left(\frac{R + u}{\sigma_{\psi}}\right) - \tilde{\mu} - R\right] \quad \text{(W3)}$$

where $\tilde{\mu} = \mu + \xi$ and was used that $\epsilon \sim \mathcal{N}(0; \sigma_{\epsilon})$

Next, we plug-in the distribution from Equation W3 in Equation W2 and obtain:

$$c = \int_{\sigma_{\psi} T^{-1}\left(\frac{R+z}{\sigma_{\psi}}\right) - \tilde{\mu} - R}^{\infty} \left(\sigma_{\psi} \cdot T\left(\frac{\tilde{\mu} + t + R}{\sigma_{\psi}}\right) - R - z\right) d\Phi(t)$$
$$= \sigma_{\psi} \int_{\theta}^{\infty} T\left(\frac{\tilde{\mu} + R + t}{\sigma_{\psi}}\right) - T\left(\frac{\tilde{\mu} + R + \theta}{\sigma_{\psi}}\right) d\Phi(t) \quad \text{(W4)}$$

where we used the substitution $z = T\left(\frac{\tilde{\mu}+R+\theta}{\sigma_\psi}\right) - R$

Web Appendix E: Approximating the Solution to the Equation

In the paper, we made an assumption that $\sigma_\epsilon = 1$. Thus, from Equation W4 it could be seen that the reservation utility is a function of three parameters: $z^* = f(\tilde{\mu} + R, \sigma_\psi, c) = f(x_1, x_2, x_3)$. During the optimization algorithm – finding this function for each customer-product combination is not feasible as it involves many integration steps.

To circumvent the computational burden, we used the trilinear interpolation technique. Specifically, for three-dimensional variables (x_1, x_2, x_3), we constructed a grid of values and computed the exact reservation utilities for each element of the grid. Notice that in this case, the space of possible values of (x_1, x_2, x_3) is divided into 3-dimensional cubes. For each of these cubes, we know the exact values of reservation utilities in eight vertices. For any vector within the cube, we approximate the reservation utility function $f(x_1, x_2, x_3)$ as:

$$\begin{gathered} f_{true}(x_1, x_2, x_3) \simeq f_{approx}(x_1, x_2, x_3) \\ = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 x_2 + \alpha_5 x_2 x_3 + \alpha_6 x_1 x_3 + \alpha_7 x_1 x_2 x_3 \end{gathered} \tag{W5}$$

where we require $f_{true}(x_1, x_2, x_3) = f_{approx}(x_1, x_2, x_3)$ at the grid (or cube vertices) points. Because $f_{approx}(x_1, x_2, x_3)$ has eight parameters and eight constraints, the linear system has a unique solution for each cell.

Web Appendix F: Derivation of Equivalent Set of Constraints on Model Parameters

After combining Equations (4-7), we can compute the variable W_i. For compactness and without loss of generality, we drop the customer index i:

$$\begin{aligned} W = &\mathbb{I}\left[\omega_b \geq \max_{s=0..C} \omega_s\right] \mathbb{I}[\mu_b + \xi_b + \epsilon_b + \psi_b \leq -R] \\ &\times \prod_{j=0}^{C-1} \left[\mathbb{I}\left[z_{j+1} \geq \max_{s=j+2..V} z_s\right] \mathbb{I}\left[\max_{s=0..j} \omega_s < \max_{s=j+1..V} z_s\right]\right] \\ &\times \mathbb{I}\left[\max_{s=0..C} \omega_s \geq \max_{s=C+1..V} z_s\right] \end{aligned} \tag{W6}$$

Consider the part of the equation:

$$\begin{aligned} P_1 &= \prod_{j=0}^{C-1} \left[\mathbb{I}\left[z_{j+1} \geq \max_{s=j+2..V} z_s\right]\right] \\ &= \prod_{j=0}^{C-1} \left[\prod_{s=j+2}^{V} \mathbb{I}[z_{j+1} \geq z_s]\right] \\ &= \mathbb{I}\left[z_C \geq \max_{s=C+1..V} z_s\right] \prod_{j=1}^{C-1} \mathbb{I}[z_j \geq z_{j+1}] \end{aligned} \tag{W7}$$

Notice that Equation W7 is a necessary condition for $W = 1$. Thus, we can assume that these inequalities hold in further derivations. Specifically, it follows that $\max_{s=j+1..V} z_s = z_{j+1}$ for $j \leq C$ and we can rewrite another part of the equation as:

$$\begin{aligned} P_2 &= \prod_{j=0}^{C-1} \mathbb{I}\left[\max_{s=0..j} \omega_s < \max_{s=j+1..V} z_s\right] = \prod_{j=0}^{C-1} \mathbb{I}\left[\max_{s=0..j} \omega_s < z_{j+1}\right] \\ &= \prod_{j=0}^{C-1} \prod_{s=0}^{j} \mathbb{I}[\omega_s < z_{j+1}] = \prod_{j=0}^{C-1} \mathbb{I}[\omega_j < z_C] \end{aligned} \tag{W8}$$

Similarly, we find:

$$\begin{aligned} P_3 &= \mathbb{I}\left[\max_{s=0..C} \omega_s \geq \max_{s=C+1..V} z_s\right] \mathbb{I}\left[\omega_b \geq \max_{s=0..C} \omega_s\right] \\ &= \mathbb{I}\left[\omega_b \geq \max_{s=C+1..V} z_s\right] \prod_{j=0}^{C} \mathbb{I}[\omega_j < \omega_b] \end{aligned} \tag{W9}$$

Finally, after combining all equation for P_1, P_2 and P_3 we obtain the simplified version of variable W:

$$W = \left[\prod_{j=1}^{C-1} \mathbb{I}\left[z_j \geq z_{j+1}\right]\right] \mathbb{I}\left[z_C \geq \max_{s=C+1..V} z_s\right] \left[\prod_{j=0}^{C-1} \mathbb{I}\left[\omega_j \leq \min\{z_C, \omega_b\}\right]\right] \\ \mathbb{I}[\omega_C \leq \omega_b] \mathbb{I}\left[\omega_b \geq \max_{s=C+1..V} z_s\right] \mathbb{I}[\mu_b + \xi_b + \epsilon_b + \psi_b \leq -R] \tag{W10}$$

Web Appendix G: Derivation of Semi-Closed Form Likelihood

As in the previous sections, we drop the customer-related index i for compactness. Recall the set of constraints that must be satisfied in order to observe a given customer sequence.

$$W = \left[\prod_{j=1}^{C-1} \mathbb{I}\left[z_j \geq z_{j+1}\right]\right] \mathbb{I}\left[z_C \geq \max_{s=C+1..V} z_s\right] \left[\prod_{j=0}^{C-1} \mathbb{I}\left[\omega_j \leq \min\{z_C, \omega_b\}\right]\right] \\ \mathbb{I}[\omega_C \leq \omega_b] \mathbb{I}\left[\omega_b \geq \max_{s=C+1..V} z_s\right] \mathbb{I}[\mu_b + \xi_b + \epsilon_b + \psi_b \leq -R] = 1 \tag{W11}$$

where ω_j is a function of unobserved to researcher shocks ξ_j and ϵ_j; z_j is a function of unobserved to researcher shock ξ_j. Because we assumed that all shocks are independent, we can rewrite the probability as:

$$\iiint W dF(\boldsymbol{\xi}, \boldsymbol{\epsilon}, \boldsymbol{\psi}) = \iiint W \prod_{j=1}^{V} dF_{\xi_j}(\xi_j) \prod_{j=1}^{C} dF_{\epsilon_j}(\epsilon_j)\, dF_{\psi_b}(\psi_b) \tag{W12}$$

The distribution $F_{\psi_b}(\psi_b)$ is known and because for independent variables holds $F(\psi) = F(\psi|\epsilon)$ we can integrate out the variable ψ_b as only one constraint depends on it:

$$\int \mathbb{I}[\mu_b + \xi_b + \epsilon_b + \psi_b \leq -R]\, dF_{\psi_b}(\psi_b) = 1 - \Phi\left[-\frac{R + \mu_b + \xi_b + \epsilon_b}{\sigma_{\psi_b}}\right] \tag{W13}$$

Next, $\xi_j: j = C + 1 \ldots V$ appears only in two constraints that could be simplified to:

$$\begin{aligned}
&\iiint \mathbb{I}\left[z_C \geq \max_{s=C+1..V} z_s\right] \mathbb{I}\left[\omega_b \geq \max_{s=C+1..V} z_s\right] \prod_{j=C+1}^{V} dF_{\xi_j}(\xi_j) \\
&= \iiint \mathbb{I}\left[\min\{z_C, \omega_b\} \geq \max_{s=C+1..V} z_s\right] \prod_{j=C+1}^{V} dF_{\xi_j}(\xi_j) \\
&= \prod_{j=C+1}^{V} \int \mathbb{I}\left[\min\{z_C, \omega_b\} \geq z_j\right] dF_{\xi_j}(\xi_j) \qquad \text{(W14)} \\
&= \prod_{j=C+1}^{V} \int \mathbb{I}\left[z_j^{-1}(\min\{z_C, \omega_b\}) \geq \xi_j\right] dF_{\xi_j}(\xi_j) \\
&= \prod_{j=C+1}^{V} \left[1 - F_{\xi_j}\left(z_j^{-1}(\min\{z_C, \omega_b\})\right)\right]
\end{aligned}$$

where we used the fact that $z_j(\xi_j)$ is an invertible function for each j and the distribution ξ_j is known.

Next, we modify the constraints related to a purchase decision $\mathbb{I}[\omega_C \leq \omega_b] \prod_{j=0}^{C-1} \mathbb{I}\left[\omega_j \leq \min\{z_C, \omega_b\}\right]$. However, in this case, we need to consider three separate cases: choosing an outside option, choosing the last searched option, and all else.

Choose an outside option (or $b = 0$). Shocks $\{\epsilon_j : j = 1 \ldots C\}$ could be integrated out:

$$\begin{aligned}
&\iiint \mathbb{I}[\omega_C \leq \omega_b] \prod_{j=0}^{C-1} \mathbb{I}\left[\omega_j \leq \min\{z_C, \omega_b\}\right] \prod_{j=1, j\neq b}^{C} dF_{\epsilon_j}(\epsilon_j) \\
&= \iiint \mathbb{I}[\omega_C \leq \omega_0] \mathbb{I}[\omega_0 \leq z_C] \prod_{j=1}^{C-1} \mathbb{I}\left[\omega_j \leq \min\{z_C, \omega_0\}\right] \prod_{j=1}^{C} dF_{\epsilon_j}(\epsilon_j) \qquad \text{(W15)} \\
&= \mathbb{I}[\omega_0 \leq z_C] \iiint \mathbb{I}[\epsilon_C \leq \omega_C^{-1}(\omega_0)] \prod_{j=1}^{C-1} \mathbb{I}\left[\epsilon_j \leq \omega_j^{-1}(\min\{z_C, \omega_0\})\right] \prod_{j=1}^{C} dF_{\epsilon_j}(\epsilon_j) \\
&= \mathbb{I}[\omega_0 \leq z_C] F_{\epsilon_C}\left(\omega_C^{-1}(\omega_0)\right) \prod_{j=1}^{C-1} F_{\epsilon_j}\left(\omega_j^{-1}(\min\{z_C, \omega_0\})\right)
\end{aligned}$$

For the case when $b \neq 0$ the situation is slightly more complicated as the constraint related to product returns depends on ϵ_b. Therefore, it should be included in derivations, we denote the result in Equation W15 as $H_b(\epsilon_b)$ and proceed to the case $b \neq 0$.

Choose the last clicked option (or $b = C$).

$$\begin{aligned}
&\iiint H_b(\epsilon_b)\mathbb{I}[\omega_C \leq \omega_b]\prod_{j=0}^{C-1}\mathbb{I}\big[\omega_j \leq \min\{z_C,\omega_b\}\big]\prod_{j=1}^{C} dF_{\epsilon_j}\big(\epsilon_j\big) \\
&= \iiint H_C(\epsilon_C)\mathbb{I}[\omega_0 \leq z_C]\mathbb{I}[\omega_0 \leq \omega_C]\prod_{j=1}^{C-1}\mathbb{I}\big[\omega_j \leq \min\{z_C,\omega_C\}\big]\prod_{j=1}^{C} dF_{\epsilon_j}\big(\epsilon_j\big) \\
&= \mathbb{I}[\omega_0 \leq z_C]\int H_C(\epsilon_C)\mathbb{I}[\omega_0 \leq \omega_C]\prod_{j=1}^{C-1} F_{\epsilon_j}\left(\omega_j^{-1}(\min\{z_C,\omega_C\})\right) dF_{\epsilon_C}(\epsilon_C) \\
&= \mathbb{I}[\omega_0 \leq z_C]\int_{\omega_C^{-1}(\omega_0)}^{\infty} H_C(\epsilon_C)\prod_{j=1}^{C-1} F_{\epsilon_j}\left(\omega_j^{-1}(\min\{z_C,\omega_C\})\right) dF_{\epsilon_C}(\epsilon_C)
\end{aligned} \tag{W16}$$

Choose other option (or $0 < b < C$).

$$\begin{aligned}
&\iiint H_b(\epsilon_b)\mathbb{I}[\omega_C \leq \omega_b]\prod_{j=0}^{C-1}\mathbb{I}\big[\omega_j \leq \min\{z_C,\omega_b\}\big]\prod_{j=1}^{C} dF_{\epsilon_j}\big(\epsilon_j\big) \\
&= \iiint H_b(\epsilon_b)\mathbb{I}[\omega_C \leq \omega_b \leq z_c]\prod_{j=0, j\neq b}^{C-1}\mathbb{I}\big[\omega_j \leq \min\{z_C,\omega_b\}\big]\prod_{j=1}^{C} dF_{\epsilon_j}\big(\epsilon_j\big) \\
&= \iiint H_b(\epsilon_b)\mathbb{I}[\omega_C \leq \omega_b \leq z_C]\,\mathbb{I}[\omega_0 \leq \min\{z_C,\omega_b\}] \\
&\times \prod_{j=1, j\neq b}^{C-1}\mathbb{I}\big[\omega_j \leq \min\{z_C,\omega_b\}\big]\prod_{j=1}^{C} dF_{\epsilon_j}\big(\epsilon_j\big) \\
&= \mathbb{I}[\omega_0 \leq z_C]\int_{\omega_b^{-1}(\omega_0)}^{\omega_b^{-1}(z_C)} H_b(\epsilon_b)F_{\epsilon_C}\Big(\omega_C^{-1}(\omega_b)\Big) \\
&\times \prod_{j=1, j\neq b}^{C-1} F_{\epsilon_j}\left(\omega_j^{-1}(\min\{z_C,\omega_b\})\right) dF_{\epsilon_b}(\epsilon_b)
\end{aligned} \tag{W17}$$

After we combine the deviations above, we rewrite the original integral in the form of:

$$P = \iiint W dF(\boldsymbol{\xi}, \boldsymbol{\epsilon}, \boldsymbol{\psi}) = \iiint \prod_{j=1}^{C-1} \mathbb{I}[z_j \geq z_{j+1}]\, \mathbb{I}[\omega_0 \leq z_C] \int_{\underline{\epsilon_b}}^{\overline{\epsilon_b}} B(\xi_C, \epsilon_b) dF_{\epsilon_b}(\epsilon_b) \prod_{j=1}^{C} dF_{\xi_j}(\xi_j) \tag{W18}$$

where $B(\cdot,\cdot)$ is a function which depends on two unobserved shocks ξ_C through z_C and ϵ_b through ω_b. Note that the function B itself is formally a function of all parameters of the model, but notable it depends only on two unobservable to the researcher variables.

Next notice that only z_C depends on ξ_C, thus the previous equation could be rewritten as:

$$\begin{aligned} P &= \iiint W dF(\boldsymbol{\xi}, \boldsymbol{\epsilon}, \boldsymbol{\psi}) \\ &= \iiint \prod_{j=1}^{C-1} \mathbb{I}[z_j \geq z_{j+1}]\, \mathbb{I}[\omega_0 \leq z_C] \int_{\underline{\epsilon_b}}^{\overline{\epsilon_b}} B(\xi_C, \epsilon_b) dF_{\epsilon_b}(\epsilon_b) \prod_{j=1}^{C} dF_{\xi_j}(\xi_j) \\ &= \int_{-\infty}^{+\infty} \int_{\underline{\epsilon_b}}^{\overline{\epsilon_b}} \mathbb{I}[\omega_0 \leq z_C] B(\xi_C, \epsilon_b)\, D(\xi_C) dF_{\epsilon_b}(\epsilon_b) dF_{\xi_C}(\xi_C) \end{aligned} \tag{W19}$$

where $D(\xi_C) = \iiint \prod_{j=1}^{C-1} \mathbb{I}[z_j \geq z_{j+1}] \prod_{j=1}^{C-1} dF_{\xi_j}(\xi_j)$

Notice that integral in $D(\xi_C)$ has a chain-like structure. Thus, we can sample random shocks iteratively. Let's assume we sampled some value of ξ_C with ξ_C^g being a realization of this random variable (thus, $z_C^g = z_C(\xi_C^g)$ also sampled). In this case, we may sample ξ_{C-1} in a way that $z_{C-1} \geq z_C$ (or $\xi_{C-1} \leq z_{C-1}^{-1}\left(z_C(\xi_C^g)\right)$ for random shock itself). However, we should adjust for the probability of such event $F_{\xi_{C-1}}\left(z_{C-1}^{-1}\left(z_C(\xi_C^g)\right)\right)$. Notice that after generating ξ_{C-1}^g we can repeat this procedure for ξ_{C-2}, and so on.

After recursively applying the procedure discussed in the previous paragraph for each j we obtain the random sample $(\xi_1^g, \ldots \xi_C^g)$ such that $\prod_{j=1}^{C-1} \mathbb{I}[z_j \geq z_{j+1}] = 1$ but the probability needs to be adjusted by $\prod_{j=1}^{C-1} F_{\xi_j}\left(z_j^{-1}\left(z_{j+1}(\xi_{+1}^g)\right)\right)$.

Finally, we can eliminate the $\mathbb{I}[\omega_0 \le z_C]$ by sampling ξ_C from a distribution such that $\omega_0 \le z_C$ holds and adjust the probability by $F_{\xi_C}\left(z_C^{-1}(\omega_0)\right)$.

Web Appendix H: Details on the Model Estimation and Analysis of Synthetic Data

Modified maximization procedure. In the results section we mentioned that in the empirical part of the paper, we slightly modified the maximization procedure. Specifically, we firstly estimated the model on the data on customers who viewed up to 48 products (one product listing). We used the estimated values for β^{ψ} and R (parameters related to product returns) as fixed and re-estimated the remaining parameters on the full data. This was done to account data imbalance – in Equation 8, there is only one constraint directly related to product returns $\mathbb{I}[\mu_{ib} + \xi_{ib} + \epsilon_{ib} + \psi_{ib} \le -R_i]$. Moreover, a large portion of sessions (~95%) ended up without purchase and thus could not be used to estimate the parameters related to product returns. As a result, when optimizing the model on the full data, the impact of "return"-related constraints on the final likelihood becomes negligible and parameters related to returns are "regularized" to zero. This leads to a substantial loss of accuracy in the "return"-related parameters with long browsing sessions (many clicks).

Additional analysis on synthetic data. We discussed various estimation approaches which were used in previous research. Table W1 reports the results of the estimation. It demonstrates that methods used in the previous research are not suitable for the model discussed in the paper. In the paper we provide several reasons why old methods are failing, the detailed investigation of the best suitable method goes beyond the scope of the paper.

Table W1: Comparison to Alternative Estimation Approaches.

	True value	Approach in Paper	AR Simulator[a]	AR Simulator (with smoothing)[b]
β_0^u	– 4.40	– 4.38	– 4.05	– 3.37
β_1^u	– .30	– .31	.00	.30
σ_1^u	.50	.43	.00	.04
β_0^c	– 7.00	– 7.00	– 7.80	– 8.19
β_1^c	.50	.52	.95	.32
β_0^ψ	1.00	1.09	.00	1.02
$\log R$	– 1.00	– .93	– 1.22	2.02

[a]AR (Accept-reject) Simulator was used with 10^5 simulated random shocks.
[b]AR Simulator (with smoothing) was used with 10^3 simulated random shocks.