# Online Appendix for
# "Sources of limited consideration and market power in e-commerce"*

Michael Sullivan

Department of Economics, University of Western Ontario

October 10, 2024

## O.1 Expressions for search effort outcome probabilities

This appendix provides chains of inequalities relating indirect and reservation utilities for every possible search effort outcome in the model for an arbitrary number of retailers $F$. Let $\mathcal{F} = \{1, 2, \ldots, F\}$. I suppress consumer $i$, brand $j$, and search effort $t$ subscripts for notational brevity here as they do not play a role in the analysis. Let $\tilde{f} = \{f_1, f_2, \ldots, f_k\}$ be the sequence of stores visited by the consumer, and let $f^*$ be the store from which the consumer makes a purchase. Let $\bar{r} = \max_{f \in \mathcal{F} \setminus \tilde{f}} r_f$ be the greatest reservation utility among all non-searched stores. Let $\bar{u} = \max_{f \in \tilde{f} \setminus \{f^*\}} u_f$ be the indirect utilities of all searched stores excluding the store of ultimate purchase.

Consider first the case in which $\tilde{f}$ includes only a single store: $\tilde{f} = \{f_1\}$. The consumer's choice to visit $f_1$ and then choose the outside option corresponds to one of the following chains of inequalities:

$$\begin{aligned} r_{f_1} &\geq u_0 \geq u_{f_1} \vee \bar{r} \\ u_0 &\geq r_{f_1} \geq u_{f_1} \vee \bar{r} \\ u_0 \geq u_{f_1} &\geq r_{f_1} \geq \bar{r}. \end{aligned} \tag{1}$$

*Email address: msulli65@uwo.ca. This project draws on research supported by the Social Sciences and Humanities Research Council.

It is possible for the consumer to visit store $f_1$ when the outside option's indirect utility exceeds $f_1$'s reservation utility because, by assumption, the consumer must visit at least one store. The first chain of inequalities above corresponds to the case in which the consumer would have searched without a requirement to visit at least one store, but only wishes to visit store $f_1$. The next two chains of inequalities correspond to the case in which the consumer would not have searched without this requirement, and then finds an indirect utility that falls below that of the outside option. Under the article's maintained distributional assumptions, the probability of the first chain of inequalities is

$$\frac{e^{\bar{r}_{f_1}}}{e^{\bar{u}_0} + e^{\bar{u}_{f_1}} + \sum_{g=1}^{F} e^{\bar{r}_g}} \times \frac{e^{\bar{u}_0}}{e^{\bar{u}_0} + e^{\bar{u}_{f_1}} + \sum_{g \notin \{0, f_1\}}^{F} e^{\bar{r}_g}} \tag{2}$$

for $\bar{u}_g = u_g - \varepsilon_g$ and $\bar{r}_g = r_g - \eta_g$. The probability of the search effort outcome described above is the sum of the probabilities of the chains of inequalities in (1). I will not explicitly state any more choice probabilities, however, since they follow the same rank-order logit form as (2).

Now consider the case in which $i$ buys from $f_1$ after visiting $f_1$ alone. The inequalities corresponding to this outcome are

$$\begin{aligned} r_{f_1} &\geq u_{f_1} \geq u_0 \vee \bar{r} \\ u_{f_1} &\geq r_{f_1} \geq u_0 \vee \bar{r} \\ u_{f_1} &\geq u_0 \geq r_{f_1} \geq \bar{r}. \end{aligned}$$

Next consider the case in which $\tilde{f}$ includes more than one alternative. There are two cases to consider. The first is that in which the consumer buys from $f^* = 0$ (the outside option) or $f^* \in \tilde{f} \setminus \{f_k\}$ (any store except the last visited). The chain of inequalities corresponding to this case is

$$r_{f_1} \geq r_{f_2} \geq \cdots \geq r_{f_{k-1}} \geq r_{f_k} \geq u_{f^*} \geq \bar{u} \vee \bar{r}.$$

Here, the fact that $r_{f_j} \geq r_{f_{j+1}}$ for $j \in \{1, \ldots, k-1\}$ is a necessary and sufficient condition for $f_j$ being visited before $f_{j+1}$ under the optimal sequential search strategy (conditioning on the consumer visiting both stores). The fact that $r_{f_k} \geq u_{f^*} \geq \bar{u} \vee \bar{r}$ is a necessary and sufficient condition for the consumer to terminate search after visiting

$f_k$ and to purchase from $f^*$. The fact that $r_{f_j} \geq u_{f^*}$ for all $j \leq k$ ensures that the consumer does not terminate search until visiting store $f_k$. Last, $u_{f^*} \geq \bar{u}$ ensures that the consumer purchases from $f^*$ rather than from any other seller.

The second case to consider is that in which the consumer purchases from $f_k$, the final store visited. Above, it was not possible for the indirect utility $u_{f_j}$ of seller $f_j$ with $j < k$ to exceed the reservation utility of any visited store: in that case, the consumer would have stopped to purchase the good rather than continue searching until $f_k$. In the second case described above, though, this is possible; as an implication, there are multiple chains of inequalities corresponding to outcomes covered by the second case. These chains of inequalities are

$$\begin{aligned}
&r_{f_1} \geq r_{f_2} \geq \cdots \geq r_{f_{k-1}} \geq r_{f_k} \geq u_{f_k} \geq \bar{u} \vee \bar{r} \\
&r_{f_1} \geq r_{f_2} \geq \cdots \geq r_{f_{k-1}} \geq u_{f_k} \geq r_{f_k} \geq \bar{u} \vee \bar{r} \\
&r_{f_1} \geq r_{f_2} \geq \cdots \geq u_{f_k} \geq r_{f_{k-1}} \geq r_{f_k} \geq \bar{u} \vee \bar{r} \\
&\qquad\qquad\vdots \\
&r_{f_1} \geq u_{f_k} \geq \cdots \geq r_{f_{k-1}} \geq r_{f_k} \geq \bar{u} \vee \bar{r} \\
&u_{f_k} \geq r_{f_1} \geq \cdots \geq r_{f_{k-1}} \geq r_{f_k} \geq \bar{u} \vee \bar{r}.
\end{aligned}$$

An alternative and perhaps simpler characterization of these inequalities is

$$r_{f_1} \geq r_{f_2} \geq \cdots \geq r_{f_{k-1}} \geq r_{f_k} \geq \bar{u} \vee \bar{r} \qquad \text{AND} \qquad u_{f_k} \geq \bar{u} \vee \bar{r}.$$

The first chain of inequalities provides a sufficient and necessary condition for the consumer to search until $f_k$ in the order given by $\tilde{f}$. The second inequality (after "AND") is a necessary and sufficient condition for the consumer to purchase from $f_k$ rather than any other retailer and to terminate search after visiting $f_k$.

## O.2 Construction of search effort panel

In constructing a search effort around a transaction, I include all visits to 1800 or VM in the $K$ days before the transaction and all visits to WM in the $K' \leq K$ days before. In the baseline specification, $K = 14$ and $K' = 2$. I consider alternative values in Section 3. The reason for using a shorter time window for WM is that consumers may visit Walmart for purposes unrelated to contact lenses; a shorter window may exclude such visits. I also construct a search effort for each visit to 1800 or VM that does not result in a transaction. In doing so, I search for visits to retailers within $R$ days (1800 and VD) or $R'$ days (WM) of this visit, and I assign these visits to the search effort of the initial visit. In the baseline specification, $R = 7$ and $R' = 2$. I proceed to add visits that are within $R$ (1800 and VM) or $R'$ (WM) days of visits that have already been added to the search effort, and I continue to iteratively add visits until no more visits are added.

## O.3 Structure of regressions underlying the I-I estimator

Let $Y_n = \{y_i\}_{i=1}^n$ denote the collection of search effort outcomes in the estimation sample, where $y_i = \{y_{it}\}_{t=1}^{T_i}$ and $y_{it}$ is a vector of search outcomes for consumer $i$ in search effort $t$ (i.e., the sequence of stores that consumer $i$ visited in search effort $t$ and consumer $i$'s purchase decision in search effort $t$). Next, let $X_n = \{x_i\}_{i=1}^n$ denote the collection of explanatory variables in the estimation sample, where $x_i = \{x_{it}\}_{t=1}^{T_i}$ and $x_{it}$ is a vector including the prices for consumer $i$'s prescribed brand of contact lenses during search effort $t$ as well as the consumer's state during search effort $t$.[1] The statistic $\hat{\beta}_n$ is the value of $\beta$ minimizing the criterion function

$$Q_n(Y_n, X_n, \beta) = \frac{1}{n} \sum_{i=1}^{n} g(y_i, x_i, \beta).$$

where

$$g(y_i, x_i, \beta) = \sum_{j=1}^{J} \sum_{t=1}^{T_i} w_{ijt}(y_{it,j} - x'_{it,j}\beta_j)^2.$$

Note that the value of $\beta$ minimizing the criterion function is the vector obtained by stacking $J$ weighted least squares estimators, each computed on a dataset of search efforts. Each $j$ corresponds to a distinct regression with a dependent variable $y_{it,j}$ computed from $y_{it}$. Similarly, each $x_{it,j}$ is some vector-valued transformation of $x_{it}$ that is used as the regressor vector in the $j$th regression. The weights $w_{ijt}$ will generally depend on the data $(y_i, x_i)$. Consider, e.g., the regression $j$ corresponding to the share of search efforts in which a consumer in state $h_{ift} = 1$ visits store $g$. In this case, $y_{it,j}$ is an indicator for whether consumer $i$ visited store $g$ in search effort $f$, $x_{it,j} = 1$, and $w_{ijt}$ is an indicator for whether consumer $i$'s state at search effort $t$ was $h_{ift} = 1$.

[1] This is a minor abuse of notation, since I use $y_i$ and $x_i$ to signify subtly different random elements in the main structural model and in the auxiliary model.

The auxiliary model statistics computed on data that are simulated under structural model parameter $\theta$ are defined by

$$\tilde{\beta}_n^H(\theta) = \arg\min_{\beta \in B} Q_{nH}(\tilde{Y}_n^H(\theta), \tilde{X}_n^H, \beta).$$

Here, $H$ is the number of simulates, $\tilde{Y}_n^H(\theta)$ are outcome variables simulated under $\theta$ conditional on $\tilde{X}_n^H$, and $\tilde{X}_n^H$ is constructed by repeating $X_n$ $H$ times.

## O.4 Optimal weighting matrix and inference

The asymptotic normality of the I-I estimator is ensured by conditions that are standard in the I-I literature.[2] The asymptotic normality result for the I-I estimator $\hat{\theta}_n^H(\Omega)$ is

$$\sqrt{n}(\hat{\theta}_n^H(\Omega) - \theta_0) \to_d N\left(0, V_{\hat{\theta}_n^H}(\Omega)\right), \; V_{\hat{\theta}_n^H}(\Omega) = (B_0'\Omega B_0)^{-1} B_0'\Omega\Gamma_0^{-1} V_{\hat{\beta}} \Gamma_0^{-1}\Omega B_0 (B_0'\Omega B_0)^{-1}$$

where

$$\begin{aligned} V_{\hat{\beta}} &= \mathrm{Var}\left(s_{i0} - \frac{1}{H}\sum_{h=1}^{H} s_{ih}\right), s_{ih} = \begin{cases} \frac{\partial g}{\partial \beta}(y_i, x_i, \beta_0), & h = 0, \\ \frac{\partial g}{\partial \beta}(\tilde{y}_i^h(\theta_0), x_i, \beta_0), & h \in \{1, \ldots, H\} \end{cases} \\ \Gamma_0 &= \frac{\partial^2 Q}{\partial \beta \partial \beta'}(\beta_0; \theta_0) \\ B_0 &= \frac{\partial b}{\partial \theta}(\theta_0). \end{aligned}$$

In the definitions above, $\tilde{y}_i^h(\theta_0)$ are search effort outcomes simulated under model parameters $\theta_0$ and $Q(\beta;\theta)$ is the population criterion function, i.e., the uniform probability limit of $Q_n(Y_n, X_n, \beta)$ as $n \to \infty$ when $(Y_n, X_n)$ are generated under the model with structural parameter $\theta$. Also, the binding function

$$b(\theta) = \arg\min_{\beta \in B} Q(\beta;\theta)$$

is the probability limit of the $\hat{\beta}_n$ parameters under a given vector of structural parameters $\theta$. Last, $\beta_0 = b(\theta_0)$.

[2] See Gouriéroux et al. (1993) for details.

The optimal weighting matrix $\Omega^*$ is

$$\Omega^* = \Gamma_0 V_{\hat{\beta}}^{-1} \Gamma_0.$$

A practical problem arises in calculating an empirical analogue of the optimal weighting matrix when $V_{\hat{\beta}}$ is singular or close to singular. Due to this problem, I instead use an approximately optimal weighting matrix

$$\Omega^{\text{approx}} = \Gamma_0 \left( V_{\hat{\beta}} + \epsilon I \right)^{-1} \Gamma_0,$$

where $\epsilon$ is a small number and $I$ is the identity matrix. Note that this approximation does not affect the estimator's consistency.

I estimate the weighting matrix $\Omega^{\text{approx}}$ and asymptotic variance $V_{\hat{\theta}_n^H}(\Omega)$ by replacing population objects appearing in expressions above with their sample analogues. Estimation of $B_0$ follows from estimation of $\Gamma_0$ and

$$\Lambda_0 = \Gamma_0 B_0 = -\frac{\partial^2 Q}{\partial \beta \partial \theta'}(\beta_0; \theta_0),$$

which implies $B_0 = \Gamma_0^{-1} \Lambda_0$. The estimates $\hat{\Gamma}_0$ and $\hat{\Lambda}_0$ of $\Gamma_0$ and $\Lambda_0$, respectively, are based on second- and cross-partial numerical derivatives of $Q_n$ with respect to $\beta$ and $\theta$ as evaluated at $(\hat{\beta}_n, \hat{\theta})$. My estimate $\hat{V}_{\hat{\beta}}$ of $V_{\hat{\beta}}$ is the sample variance of $s_{i0} - (1/H) \sum_h s_{ih}$ as evaluated at $(\hat{\beta}, \hat{\theta})$. Estimating the weighting matrix $\Omega^{\text{approx}}$ requires a preliminary consistent estimate $\hat{\theta}$ of $\theta_0$; I use an estimate $\hat{\theta}$ obtained upon setting the weighting matrix equal to the identity matrix. When I later estimate $\Gamma_0$, $\Lambda_0$, and $V_{\hat{\beta}}$ as inputs in the estimation of the asymptotic variance $V_{\hat{\theta}_n^H}(\Omega^{\text{approx}})$, I set $\hat{\theta}$ equal to the estimate obtained under my estimate $\hat{\Omega}^{\text{approx}}$ of $\Omega^{\text{approx}}$ (see below). The estimate of $B_0$ that I use in estimating the asymptotic variance $\hat{V}_{\hat{\theta}_n^H}(\Omega^{\text{approx}})$ is $\hat{B}_0 = \hat{\Gamma}_0^{-1} \hat{\Lambda}_0$. Last, I obtain estimates of $\Omega^{\text{approx}}$ and $V_{\hat{\theta}_n^H}(\Omega^{\text{approx}})$ by substituting the estimates of the $\Gamma_0$, $B_0$, and $V_{\hat{\beta}}$ matrices described above in for their true values:

$$\hat{\Omega}^{\text{approx}} = \hat{\Gamma}_0 \left( \hat{V}_{\hat{\beta}} + \epsilon I \right)^{-1} \hat{\Gamma}_0$$

$$\hat{V}_{\hat{\theta}_n^H}(\Omega^{\text{approx}}) = (\hat{B}_0' \hat{\Omega}^{\text{approx}} \hat{B}_0)^{-1} \hat{B}_0' \hat{\Omega}^{\text{approx}} \hat{\Gamma}_0^{-1} \hat{V}_{\hat{\beta}} \hat{\Gamma}_0^{-1} \hat{\Omega}^{\text{approx}} \hat{B}_0 (\hat{B}_0' \hat{\Omega}^{\text{approx}} \hat{B}_0)^{-1}$$

## O.5 Conditional dependence of store tastes and prices

Consider a consumer $i$ who makes multiple search efforts across time and whose initial observed purchase is from store $f_i^{\text{initial}}$. I expect that, conditional on $f_i^{\text{initial}}$, the prices that consumer $i$ faces and the consumer's unobserved tastes $\gamma_i$ will be correlated. This is because a consumer who buys from a store $f_i^{\text{initial}}$ despite its high prices requires strong positive tastes for store $f_i^{\text{initial}}$ to rationalize buying from $f_i^{\text{initial}}$.

To assess this conditional correlation, I regress an indicator for whether a consumer visited a store other than $f_i^{\text{initial}}$ on the relative price of $f_i^{\text{initial}}$ at the time of the purchase that determined the consumer's state. The regression equation is

$$\mathbb{1}\{i \text{ visits store other than } f_i^{\text{initial}} \text{ in } t\} = \lambda_0 + \lambda_1 \left( {}^{p_{j,f_i^{\text{initial}},1}}/_{\bar{p}_{j1}} \right) + \epsilon_{it}$$

where $j$ is consumer $i$'s brand; $p_{j,f_i^{\text{initial}},1}$ is the price of $f_i^{\text{initial}}$ when $i$ first purchased lenses in the sample; and $\bar{p}_{j1}$ is the mean price of $j$ across 1800, WM, and VD at the time $i$ first purchased lenses in the sample. I run the regression on a dataset including all search efforts excluding those of consumers' first purchases. A negative $\lambda_1$ estimate would indicate that consumers who bought from a relatively expensive store are less likely to consider other stores, which would indicate that these consumers have strong preferences for the store from which they historically bought contact lenses. Appendix Table O.1 provides the results. As expected, the estimate of $\lambda_1$ is negative.

Table O.1: Results for regression assessing conditional dependence of prices and store tastes

| Parameter | Estimate | SE |
|---|---|---|
| Intercept | 0.434 | 0.112 |
| Slope | -0.227 | 0.109 |

Notes: the "SE" column provides asymptotic standard errors.

## O.6 Dynamic pricing model

In addition to the analysis of static pricing in the main text, I additionally study online retailers' pricing in a dynamic framework. My approach to studying dynamic pricing in a setting with state dependence follows that of Dubé et al. (2009), who provide additional information on the properties of the general dynamic pricing model that their paper proposes and that I amend to my setting in this paper.

I analyze a model of online retailers' dynamic pricing using a Markov perfect equilibrium (MPE) solution concept. In the MPE that I consider, firms' pricing strategies maximize their payoffs subject to the constraint that their strategies condition only on information relevant to contemporaneous payoffs. This information includes the share of consumers with each value of heterogeneity $\zeta_i = \{\boldsymbol{\gamma}_i, \alpha_i\}$ that belong to each state (i.e., whose previous purchase was from each seller). It is not computationally feasible to find an MPE in a setting in which $\boldsymbol{\gamma}_i$ is continuously distributed; therefore, I compute MPE in a simplified version of the model in which $\boldsymbol{\gamma}_i$ takes on one of $K$ support points in $\mathcal{G}$. The set of types $\zeta_i$ is

$$\mathcal{Z} = \{(\boldsymbol{\gamma}, \alpha) : \boldsymbol{\gamma} \in \mathcal{G}, \alpha \in \{\alpha_0, \alpha_0 + \alpha_1\}\}.$$

Recall that $\alpha_0$ is the price sensitivity of low-income consumers and that $\alpha_1$ is the price sensitivity of high-income consumers. Let $x_{f\tau}(\zeta)$ denote the share of consumers of type $\zeta \in \mathcal{Z}$ whose previous purchase in time $\tau$ was made at store $f$, let $\mathcal{F}$ be the collection of all competing online retailers, and let $x_\tau = \{x_{f\tau}(\zeta) : f \in \mathcal{F}, \zeta \in \mathcal{Z}\}$. Following the standard terminology used in dynamic programming, I refer to $x_\tau$ as the *state* at risk of causing confusion with the consumer's state $h_i$ as defined in Section 4.

Firm $f$'s payoffs in my dynamic pricing model are the firm's present discounted profits. When players use strategies $p^* : x_\tau \mapsto p_f$, these payoffs are

$$\sum_{\tau=0}^{\infty} \beta^\tau \sum_{\zeta \in \mathcal{Z}} \mu(\zeta) \sum_g x_g(\zeta) \sigma_{fg}(p^*(x_\tau), \zeta)(p_f^*(x_\tau) - mc),$$

where $\beta$ is a discount factor shared by all competing firms, $\mu(\zeta)$ is the share of consumers of type $\zeta$, and $mc$ is firm $f$'s marginal cost of providing a consumer with a box of contact lenses. I assume that firms share a marginal cost $mc$.

The Bellman equation associated with firm $f$'s dynamic programming problem is

$$V_f(x) = \max_{p_f \geq 0} \left[ \sum_{\zeta \in \mathcal{Z}} \mu(\gamma) \sum_g x_g(\zeta) \sigma_{fg}(p_f, p^*_{-f}(x), \zeta)(p_f - mc) + \beta V_f(Q(x, p_f, p^*_{-f}(x))) \right]. \tag{3}$$

The function $Q$ appearing in (3) is the state transition function, which provides the next period's state given the contemporary state $x$ and prices $p$. The state transition is deterministically determined by consumer choice probabilities conditional on type

$\zeta_i$, state $h_i$, and prices $p$. A MPE is a pricing strategy function $p^* : x \mapsto p$ and an associated value function $V_f$ for each firm $f$ that solves the Bellman equation (3).

Implementation. To limit the size of the state space of the dynamic programming problem that I solve in finding equilibria, I drop Walmart from the analysis. Also, solving for a dynamic pricing equilibrium requires a finitely supported distribution of unobserved heterogeneity $\boldsymbol{\gamma}_i$, a marginal cost $mc$, and a discount factor $\beta$. To obtain a finitely supported distribution of $\boldsymbol{\gamma}_i$, I follow Dubé et al. (2009) in clustering consumers into a finite number of types. My clustering procedure involves (i) taking 2000 draws from my estimated unconditional distribution of $\boldsymbol{\gamma}_i$ and (ii) performing $K$-means clustering on these draws. I use the cluster centroids as the members of $\boldsymbol{\gamma}_i$'s support, and I use the share of observations in each cluster times the share of consumers with price sensitivity $\alpha$ as the corresponding population shares $\mu(\boldsymbol{\gamma}, \alpha)$ for support points $\boldsymbol{\gamma}$. Additionally, I use $K = 2$ clusters. I use information from 1-800 Contacts's quarterly report for the second quarter of 2007 to obtain a marginal cost $mc$. In particular, I divide the price of Acuvue Toric—which is the brand on which I focus in my analysis of online retailers' pricing—at 1800 in the first week of 2007 by the ratio of net sales to costs of goods and services (COGS) for January 1–June 30, 2007 as reported on 1800's quarterly report.[3] This approach applies 1800 overall markup ratio as defined in the preceding paragraph to a particular product's price to obtain an estimate of that product's marginal cost. Last, I set the discount factor $\beta$ to 0.95.

Results. Table O.2 provides percentage changes in steady-state markups under counterfactual preferences. Following Dubé et al. (2009), I compute steady-state markups by simulating an equilibrium price path from an arbitrary initial state until firms' prices converge. The initial state that I use is one in which no consumers are loyal to any online store. These results reported by Table O.2 largely accord with those obtained using a static pricing model: equilibrium markups are largely unaffected by a reduction in search costs, but markedly decrease upon an elimination of persistent unobserved heterogeneity that horizontally differentiates sellers and, to a lesser extent, upon an elimination of state dependence.

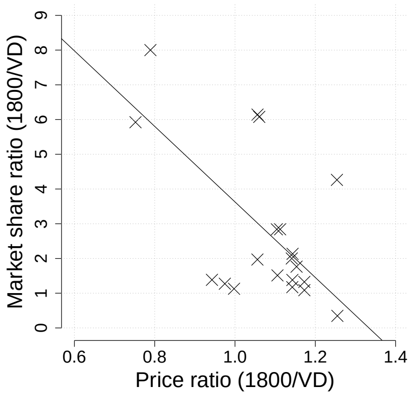
[3]Net sales and COGS were \$125,202,000 and \$73,962,000, respectively, in this time period. The ratio of these values is 1.69.

Table O.2: Percentage changes in markups from dynamic pricing model

| Store | Low search costs | No state dependence | No persistent unobs. | No price insensitive |
|---|---|---|---|---|
| 1800 | -0.5 | -0.5 | -15.9 | -4.9 |
| VD | -1.2 | -3.1 | -31.5 | -6.2 |

## O.7 Additional results

Figure O.1: Prices and intrabrand market shares at 1800 and VD




Note: Each point represents a brand. "Market share ratio (1800/VD)" provides the ratio of transactions at 1800 to those at VD. "Price ratio (1800/VD)" provides the average daily price of a brand at 1800 divided by the analogous quantity for VD. The plot includes the 20 best-selling brands and displays a least-squares line of best fit.

Table O.3: Explanatory power of consumer characteristics

| Specification | $R^2_{\text{McF}}$ | $R^2_{\text{pseudo}}$ |
|---|---|---|
| Demographic variables | 0.16 | 0.15 |
| Demographic and web use variables | 0.23 | 0.21 |

Note: this table reports measures of fit from multinomial logistic regressions in which retailer of purchase (1800, WM, or VD) is the outcome and consumer characteristics are the regressors. For the first row of the table, the included consumer characteristics are indicators for consumer having a university degree (including for a missing value for educational attainment); age groups (30s, 40s, and 50+); household income ($25-75k and over $75k); racial groups (black, Asian, and other non-white); household size (2, 3, and 4 or greater); presence of children in the household; broadband internet; Hispanic ethnicity; and census region of residence. The regression also includes year and brand fixed effects. For the second row of the table, variables characterizing consumer internet usage are also included. Each of these variables equals the number of internet browsing sessions in which the consumer visited a website in a particular category. The categories are adult, advert, career, finance, gaming, government, information, malware, media, portal, retail, social media, video, weather, web service, dating, internet/wireless companies, news, sports, travel, downloads, and directories. See Saruya and Sullivan (2023) for details. $R^2_{\text{McF}}$ reports values of McFadden's $R^2$ for the regressions, whereas $R^2_{\text{pseudo}}$ is

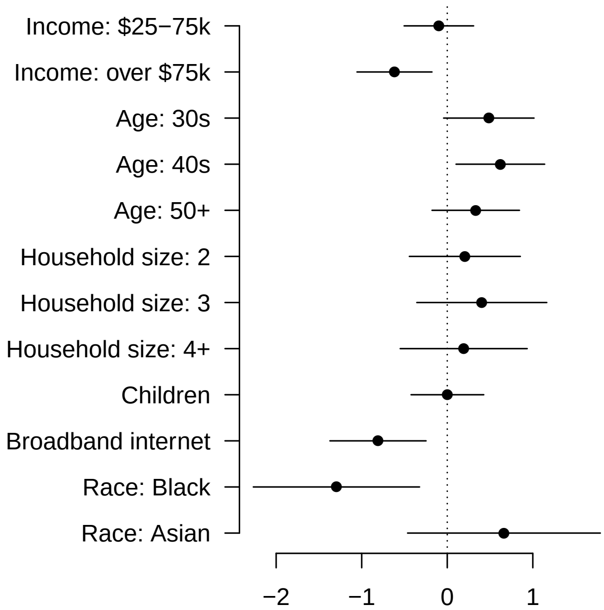
$$R^2_{\text{pseudo}} = 1 - \frac{\sum_i \sum_f (y_{if} - \hat{y}_{if})^2}{\sum_i \sum_f (y_{if} - \bar{y}_f)^2},$$

where $y_{if}$ is an indicator for observation $i$ involving a purchase from retailer $f$, $\hat{y}_{if}$ is the model-predicted probability of a purchase from retailer $f$, and $\bar{y}_f$ is the sample average of $y_{if}$.

## O.8 Robustness to treatment of Walmart

In the article's primary analysis, Walmart is treated differently than 1800 and VD, the retailers that primarily sell contact lenses: as described by Appendix O.2, I require a

Figure O.2: Selected estimates from regression of purchase choice on consumer characteristics




Notes: this figure reports selected estimates and their corresponding 95% confidence intervals from the regression described in the notes of Table O.3 that includes the web use variables.

Table O.4: Elasticity estimates for Acuvue Toric

(a) Point estimates

| | Price | | |
|---|---|---|---|
| Share | 1800 | WM | VD |
| 1800 | -2.14 | 0.18 | 0.08 |
| WM | 4.41 | -9.69 | 0.78 |
| VD | 0.23 | 0.10 | -2.14 |

(b) Standard errors

| | Price | | |
|---|---|---|---|
| Share | 1800 | WM | VD |
| 1800 | 0.66 | 0.07 | 0.03 |
| WM | 1.29 | 2.52 | 0.57 |
| VD | 0.07 | 0.05 | 0.35 |

Note: Each entry corresponds to the elasticity of long-run demand at the store indicated by the entry's row with respect to the price indicated by the entry's column. Standard errors computed using the parametric bootstrap with 100 replicates.

Table O.5: Model fit and counterfactual search patterns

| Spec. | Share visiting one store only | Mean # of visits | Share buying from... any | 1800 | VD | Share paying > min. price | Mean over-payment ($) |
|---|---|---|---|---|---|---|---|
| Observed | 0.82 | 1.20 | 0.61 | 0.36 | 0.22 | 0.66 | 3.95 |
| | - | - | - | - | - | - | - |
| Baseline | 0.84 | 1.18 | 0.55 | 0.34 | 0.20 | 0.67 | 4.13 |
| | (0.02) | (0.02) | (0.02) | (0.01) | (0.01) | (0.02) | (0.17) |
| Low search costs | 0.74 | 1.30 | 0.55 | 0.34 | 0.20 | 0.67 | 4.08 |
| | (0.02) | (0.03) | (0.02) | (0.01) | (0.01) | (0.02) | (0.18) |
| No state dep. | 0.80 | 1.22 | 0.55 | 0.33 | 0.19 | 0.67 | 4.09 |
| | (0.03) | (0.03) | (0.02) | (0.01) | (0.01) | (0.01) | (0.17) |
| No vertical diff. | 0.73 | 1.30 | 0.55 | 0.20 | 0.29 | 0.53 | 2.89 |
| | (0.04) | (0.05) | (0.02) | (0.01) | (0.03) | (0.03) | (0.22) |
| No horiz. diff. | 0.61 | 1.48 | 0.55 | 0.50 | 0.03 | 0.88 | 5.47 |
| | (0.05) | (0.07) | (0.02) | (0.05) | (0.05) | (0.07) | (0.65) |
| Logit only | 0.00 | 3.00 | 0.55 | 0.13 | 0.25 | 0.52 | 2.19 |
| | (0.00) | (0.00) | (0.02) | (0.01) | (0.01) | (0.02) | (0.14) |

Notes: the counterfactual preference changes considered are

(i) Low search costs: reduce $\bar{\kappa}$ so that the median search cost equals one half of the median search cost under the estimated value of $\bar{\kappa}$;

(ii) No state dependence: set $\phi = 0$;

(iii) No vertical differentiation: set mean store taste $Q_f = 0$ for each store $f$ to eliminate mean quality differences between stores;[4]

(iv) No horizontal differentiation: set $\gamma_{if} = \mathbb{E}[\gamma_{if}]$ for all consumers $i$ and retailers $f$; and

(v) Logit only: eliminate search costs, state dependence, vertical differentiation, and horizontal differentiation so that only prices and $\varepsilon_{ijft}$ shocks differentiate retailers.

Altering consumer preferences changes the probability that a consumer buys from any online store. To focus on the effects of qualitative preference changes rather than those of changes in the magnitude of tastes for e-commerce, I add a compensating constant $q^{\dagger}$ to each consumer's utility for every online store to ensure the outside good's share is constant across counterfactuals. I compute standard errors using a parametric bootstrap with 100 replicates.

Table O.6: Characteristics of consumers who switch between brands

| Coefficient | Estimate | SE |
|---|---|---|
| Intercept | 1.148 | 0.103 |
| Household size: 2 | -0.024 | 0.094 |
| Household size: 3 | -0.103 | 0.109 |
| Household size: 4 | 0.080 | 0.109 |
| Household size: 5 | -0.032 | 0.115 |
| Household size: 6 | -0.031 | 0.129 |
| Age: 30s | 0.078 | 0.075 |
| Age: 40s | 0.105 | 0.072 |
| Age: 50 and over | 0.070 | 0.071 |
| Children in HH? | -0.044 | 0.058 |
| $R^2$ | 0.023 | |

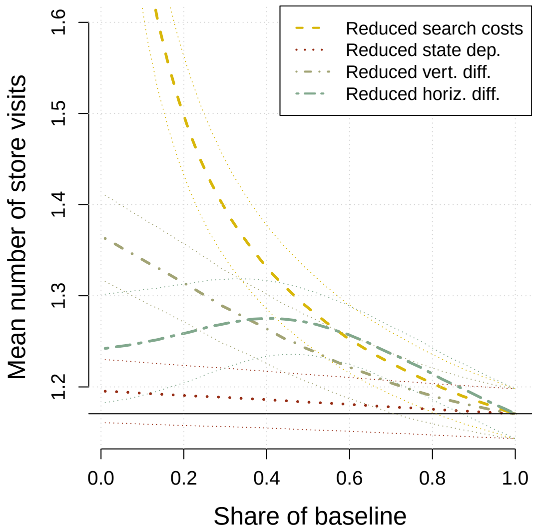
Notes: this table provides estimates from a regression of an indicator for whether a consumer ever switched brands of contact lenses on various consumer characteristics, including: (i) the size of the consumer's household, (ii) the age group of the head of the consumer's household, and (iii) an indicator for whether there is a child in the consumer's household. The number of observations is 494. The "SE" provides classical asymptotic standard errors.

Table O.7: Brand prices before and after switching

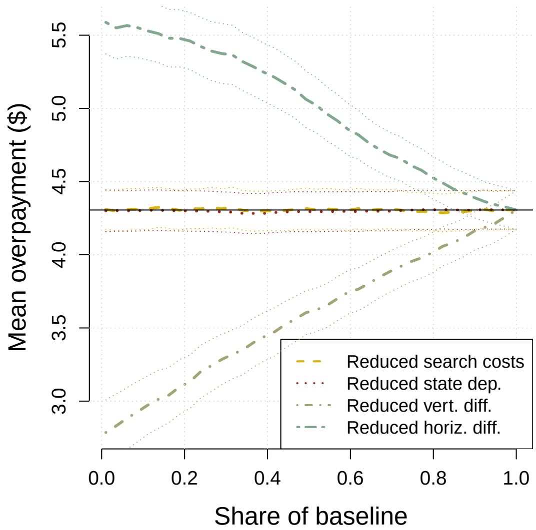
| Variable | Mean | SE |
|---|---|---|
| Before | 30.62 | 0.89 |
| After | 31.48 | 0.91 |
| Difference | 0.86 | 1.28 |
| N | 136 | - |

Notes: this table provides the mean prices of consumers' prescribed contact lens brands before and after a change of brand. In particular, it provides mean prices in a dataset of brand changes as defined as events in which the consumer purchases a brand of contact lenses different than the brand that the consumer previously purchased. The price before purchase is the mean price across retailers of the consumer's initial brand at the time of the consumer's search effort just before the change in brand. The price after purchase is the mean price across retailers of the consumer's new brand at the time of the search effort in which the consumer switches to this new brand. The number of observations is 136. The "SE" provides classical asymptotic standard errors.

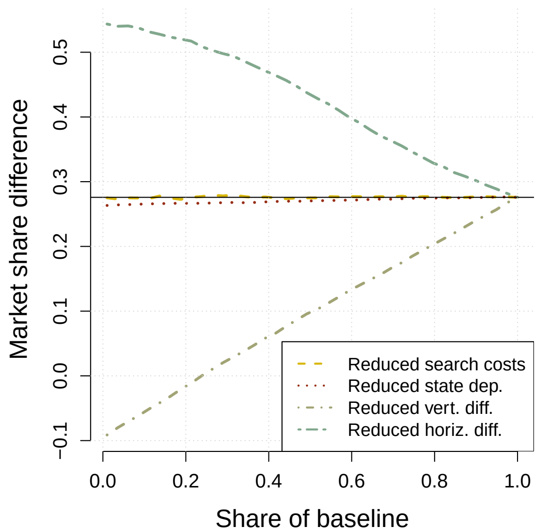
Figure O.3: Counterfactual search patterns with pointwise confidence intervals




(a) Mean number of visited retailers




(b) Mean payment over minimum available price ($)

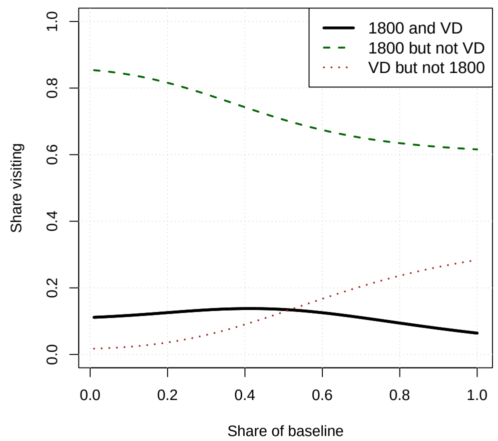



(c) Difference in market share between 1800 and VD

Notes: the dotted bands around each curve provide 95% confidence intervals as computed from a parametric bootstrap with 100 replicates.

visit to Walmart to be nearer in time to a contact lens purchase or a visit to another contact lens retailer than I require for the other retailers. In this appendix, I consider the robustness of the model estimates to instead using the same cutoffs for including a visit to Walmart in a search effort as I use for the other retailers — i.e., to setting $K' = K = 14$ and $R' = R = 2$. Changing the search effort definition in this manner

Figure O.4: Changes in mean number of visited retailers as horizontal differentiation is reduced




Notes: this figure plots (i) the share of consumers visiting both 1800 and VD; (ii) the share of of consumers visiting 1800 but not VD; and (iii) the share of consumers visiting VD but not 1800 for each extent of horizontal differentiation between zero and 100% of the baseline estimates. The main text provides details on the implementation of reductions in horizontal differentiation. Consumers visiting any of the combinations of retailers included in the figure may also possibly visit Walmart (e.g., 1800 and not VD includes consumers who visit 1800 and Walmart but not VD).

raises the share of search efforts including a visit to Walmart from 14.5% to 25.1%. In addition, I consider robustness of the parameter estimates to the inclusion of Walmart in the sample.

Table O.8 juxtaposes results obtained under the alternative cutoffs described above and obtained upon dropping Walmart from the sample to the baseline parameter estimates discussed in the main text. Each of these results is obtained using the same weighting matrix as used in computing the baseline estimates. In addition, I drop the Walmart-specific parameters and all Walmart-specific auxiliary statistics from the estimation procedure. See the notes of Table O.8 for details. The table shows that the model estimates are largely robust to the treatment of Walmart.

Table O.8: Robustness of parameter estimates to treatment of Walmart

| Parameter | Specification | | |
|---|---|---|---|
| | Baseline | Alt. cutoffs | No WM |
| $q_{1800}$ | -0.335 | -0.328 | -0.412 |
| $q_{\text{WM}}$ | -2.234 | -2.002 | - |
| $q_{\text{VD}}$ | 0.300 | 0.325 | 0.350 |
| $\phi$ | 0.493 | 0.448 | 0.493 |
| $\alpha_0$ | 0.110 | 0.107 | 0.110 |
| $\alpha_1$ | -0.084 | -0.096 | -0.063 |
| $\bar{\kappa}_{1800}$ | -2.711 | -2.762 | -3.065 |
| $\bar{\kappa}_{\text{WM}}$ | -1.887 | -2.112 | - |
| $\bar{\kappa}_{\text{VD}}$ | -1.546 | -1.504 | -1.596 |
| $\Gamma_{12}$ | -3.547 | -2.987 | - |
| $\Gamma_{13}$ | -3.257 | -3.266 | -3.257 |
| $\Gamma_{21}$ | -1.233 | -0.973 | - |
| $\Gamma_{23}$ | -1.109 | -0.852 | - |
| $\Gamma_{31}$ | -5.574 | -5.546 | -5.641 |
| $\Gamma_{32}$ | -4.097 | -2.793 | - |
| $\sigma_\gamma^2$ | 1.298 | 1.298 | 0.863 |
| $\lambda$ | 3.986 | 4.068 | 3.809 |
| Med. search cost: 1800 | 0.414 | 0.403 | 0.292 |
| Med. search cost: WM | 0.930 | 0.764 | - |
| Med. search cost: VD | 1.294 | 1.380 | 1.234 |

Notes: this table contains (i) the baseline parameter estimates described in the main text, (ii) parameter estimates obtained upon setting $K' = K = 14$ and $R' = R = 2$ ("Alt. cutoffs"; see Appendix O.2), and (iii) parameter estimates obtained upon dropping Walmart from the analysis *("No WM"). In estimating the "No WM" specification I drop the Walmart-specific parameters $q_{\text{WM}}$, $\bar{\kappa}_{\text{WM}}$, and—for all retailers $f$ other than Walmart—the parameters $\Gamma_{f,\text{WM}}$ and $\Gamma_{\text{WM},f}$. I similarly drop Walmart-specific auxiliary statistics from the estimation procedure. These are: the share of search efforts with a visit to Walmart, the intercept for Walmart in the inertia regression, all cross-visiting shares involving Walmart, and the share of search efforts with a purchase from Walmart. The table also displays median search costs (in dollars) for each retailer under each specification.

## O.9 Sensitivity of parameter estimates to auxiliary statistics

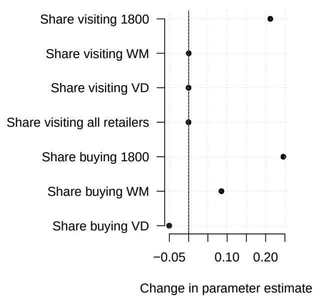
In this appendix, I assess the sensitivity of the structural model parameter estimates to the auxiliary statistics used in I-I estimation. This assessment involves re-estimating the model under alternative values of the auxiliary statistics and determining how parameter estimates obtained under these alternative values differ from the baseline parameter estimates.

First, I discuss how the availability of data on both sales and browsing behaviour permits the separate identification of search costs and parameters affecting purchase utility. Figure O.5 displays changes in the estimated retailer-specific indirect utility intercepts $q_f$ and search cost parameters $\bar{\kappa}_f$ when, for each $f \in \{1800, \text{WM}, \text{VD}\}$, (i) the share of search efforts involving a visit to $f$ increases by 5 percentage points (p.p.) and (ii) the share of search efforts involving a purchase from $f$ increases by 5 p.p. The first notable pattern in the results is that the $q_f$ indirect utility intercepts are sensitive to both the share of consumers visiting $f$ and the share of consumers purchasing from $f$, which indicates that these auxiliary statistics help pin down the estimates of the indirect utility intercepts. Additionally, the search cost parameters $\bar{\kappa}_f$ are primarily sensitive to the extent to which consumers visit $f$ rather than purchase from $f$: raising the share of search efforts involving a visit to each of 1800, WM, and VD reduces the corresponding store's $\bar{\kappa}_f$ parameter, although these parameters do not significantly fall when the store's share of purchases rises. The fact that browsing shares and purchasing shares differentially influence the $q_f$ and $\bar{\kappa}_f$ estimates reflects the argument for the separate identification of indirect utility and search cost parameters in Section 6.2 of the main text.

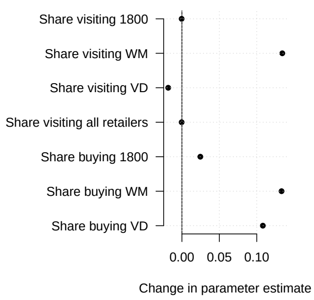
Next consider estimation of the price coefficient $\alpha_0$. Figure O.6 plots changes in the estimate of $\alpha_0$ due to changes in the values of various auxiliary statistics used in estimation; the table notes provide details on these changes. To summarize, the price coefficient is sensitive to the auxiliary statistics that are related to price sensitivity that I included to target $\alpha_0$.

I last discuss estimation of the state dependence parameter $\phi$. Figure O.7 plots changes in the estimate of $\phi$ due to changes in the values of various auxiliary statistics used in estimation. See the table notes for details. I choose the magnitude and direction of the statistic changes for interpretability and comparability. The first change is in raising the share of consumers who visit all three retailers by 1 p.p. Since state dependence reduces the extent of search, imposing more expansive search reduces the estimated extent of state dependence. Second, I reduce the share of search efforts featuring the same first-visited store as the associated consumer's previous search effort ("Inertia share"). Reducing inertia in searching naturally reduces the estimated extent of state dependence. Next, I raise the coefficients of the "role of lagged price" regressions: making consumers less sensitive to lagged prices reduces the extent of estimated state dependence, as expected. The "Inertia reg.: shift 2nd lag to 1st" row provides the change in the estimate of $\phi$ when the coefficient on the first lag of purchase in the "Inertia regression" is raised from 0.495 to 0.888 and the coefficient on the second lag is reduced from 0.392 to 0 (i.e., the first-lag coefficient is set equal to the sum of the estimated first- and second-lag coefficients whereas the second-lag coefficient is set to zero). Making consumers more sensitive to their immediately previous decisions (which drive state dependence in the model) rather than decisions further back in time (which also depend on unobserved heterogeneity) raises the estimated extent of state dependence. Last, changing the share of consumers that visit a store $f$ among those previously purchased from a store $g \neq f$ reduces the estimated extent of state dependence. This is because state dependence limits a consumer's tastes for stores other than the store from which the consumer previously made a purchase.

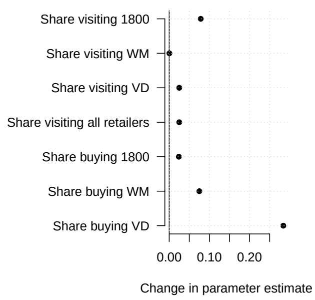
Figure O.5: Sensitivity of indirect utility and search cost parameters to auxiliary statistics



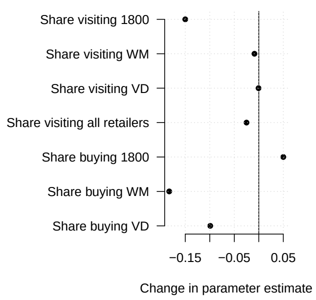

(a) $q_{1800}$



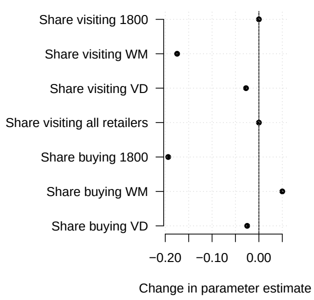

(b) $q_{\mathrm{WM}}$



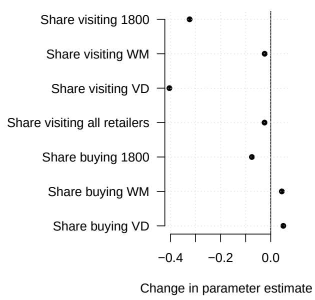

(c) $q_{\mathrm{VD}}$




(d) $\bar{\kappa}_{1800}$



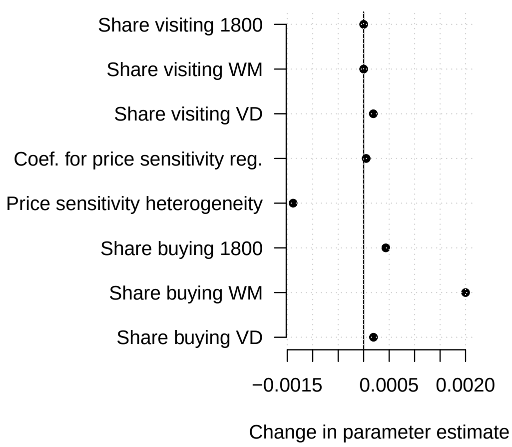

(e) $\bar{\kappa}_{\mathrm{WM}}$




(f) $\bar{\kappa}_{\mathrm{VD}}$

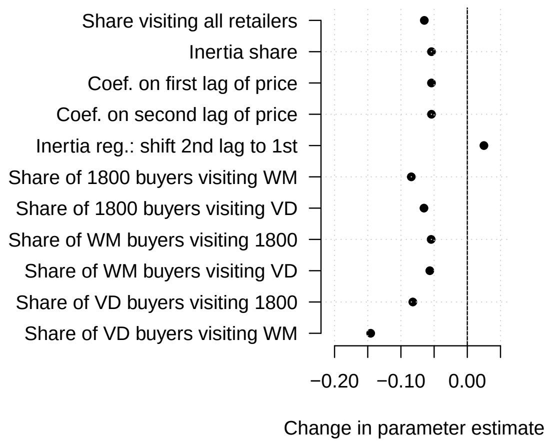
Notes: this figure provides changes in the values of the indirect utility intercept parameters $q_f$ and in the search cost parameters $\bar{\kappa}_f$ under the following changes in the I-I auxiliary statistics $\hat{\beta}_n$ used in estimation: (i) increasing the share of search efforts involving a visit to 1800 by 5 percentage points (p.p.); (ii) increasing the share of search efforts involving a visit to WM by 5 p.p.; (iii) increasing the share of search efforts involving a visit to VD by 5 p.p.; (iv) increasing the share of search efforts involving a visit to all retailers by 1 p.p.; (v) increasing the share of search efforts involving a purchase from 1800 by 5 percentage points (p.p.); (vi) increasing the share of search efforts involving a purchase from WM by 5 p.p.; and (vii) increasing the share of search efforts involving a purchase from VD by 5 p.p.

Figure O.6: Sensitivity of price coefficient $\alpha_0$ to auxiliary statistics




Notes: this figure provides changes in the esimate of the price coefficient $\alpha_0$ under the following changes in the I-I auxiliary statistics $\hat{\beta}_n$ used in estimation: (i) increasing the share of search efforts involving a visit to 1800 by 5 percentage points (p.p.); (ii) increasing the share of search efforts involving a visit to WM by 5 p.p.; (iii) increasing the share of search efforts involving a visit to VD by 5 p.p.; (iv) reducing the slope coefficient in the "price coefficient" regression described in Appendix A by 0.05, from -0.155 to -0.205 (here, I lower the slope coefficient auxiliary statistic because a lower value of this statistic corresponds to greater price sensitivity whereas a greater value of $\alpha_0$ corresponds to greater price sensitivity in the structural model); (v) raising the "price sensitivity heterogeneity" auxiliary statistic described in Appendix A by 0.1, from 0.045 to 0.145; (iv) increasing the share of search efforts involving a purchase from 1800 by 5 percentage points (p.p.); (v) increasing the share of search efforts involving a purchase from WM by 5 p.p.; and (vi) increasing the share of search efforts involving a purchase from VD by 5 p.p.

Figure O.7: Sensitivity of state dependence parameter $\phi$ to auxiliary statistics




Notes: this figure provides changes in the values of the state dependence parameter $\phi$ under the following changes in the I-I auxiliary statistics $\hat{\beta}_n$ used in estimation: (i) increasing the share of search efforts including a visit to all retailers by 1 percentage point (p.p.); (ii) reducing the inertia share (i.e., the share of search efforts with the same first-visited store as the associated consumer's previous search effort) by 10 p.p.; (iii) raising the coefficient on the first lag of price in the "role of lagged price regression by 0.5; (iv) raising the coefficient on the second lag of price in the "role of lagged price regression by 0.5; (v) setting the coefficient on the first lag of purchase in the "Inertia regression" equal to the sum of the estimated coefficients for the first and second lags and setting the coefficient on the second lag to zero; (vi) raising the share of consumers who previously bought from store $f$ that visit store $g$ in a search effort (i.e., the "cross-visiting" shares) by 5 p.p. each, one at a time.

# Bibliography

**De Los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest.** 2012. "Testing models of consumer search using data on web browsing and purchasing behavior." *American Economic Review* 102 (6): 2955–2980.

**Dubé, Jean-Pierre, Günter Hitsch, and Peter E. Rossi.** 2009. "Do switching costs make markets less competitive?" *Journal of Marketing Research* 46 (4): 435–445.

**Gouriéroux, Christian, Alain Monfort, and Eric Renault.** 1993. "Indirect inference." *Journal of Applied Econometrics* 8 (S1): S85–S118.

**Saruya, Hiroki, and Michael Sullivan.** 2023. "Cross-channel competition and complementarities in US retail." Unpublished working paper.