Consumer Search and Latent Demand

December 7 2017

Abstract

A very large fraction of online consumers typically just browse and do not make any purchase. In this paper, we estimate the latent demand from the browsing data through the lens of optimal search theory. In doing so, we demonstrate that consumer browsing histories even without choice incidence are a valuable source of information for firms to understand consumers. As an empirical application, we further gauage the potential demand that may be realized by firm side changes such as product improvement or price changes.

1 Set-up

Utility of consumer $i = 1, \cdots, I$ for option $j = 1, \cdots, J$ is

$$u_{ij} = V_{ij} + \varepsilon_{ij} + \delta_{ij},$$

where ε_{ij} is a stochastic term known to a consumer prior to search but not to the analyst. The other stochastic term δ_{ij} is the objective of search: this value is not known to consumers prior to search. Upon search (and during choice), it will be known to consumers. Analysts does not observe this quantity. We assume that ε_{ij} follows Extreme Value Type 1 distribution $\sim GEV(0, \tau, 0)$ and δ_{ij} a normal distribution$\sim N(0, \sigma^2)$. Further, assume that both random variables are i.i.d. across i and j. Lastly, consumer i's outside utility is,

$$u_{i0} \sim V_0 + \varepsilon_{io},$$

where $\varepsilon_{i0} \sim GEV(0, \tau, 0)$ and is i.i.d. across $j = 0, \cdots, J$. We assume that consumers are endowed with the known outside utility prior to search. Therefore, there is nothing for them to search for about outside option: outside utility has only one error term compared to inner options. In what follows, we supress subscript i for clarity.

2 Search order

Same as in Bart's note.

3 Choice of outside option

Let u_0 be consumer i's utility of the outside option. Let k index the options and K be the index of the last option searched. First, we partition the products: searched options are $k =< 1,2,\cdots,K >$ and they are searched in this order. Unsearched options are $m = \{K+1,\cdots,J\}$. The observation that consumer did not make any purchase means that the following must be true,

$$u_0 > u_k,\ k = 1,2,\cdots,K.$$

Note that the above provides the lower bound for the outside utility. The probability of observing the above set of events is,

$$Pr_0 = Pr(V_0 + \varepsilon_0 > V_k + \varepsilon_k + \delta_k).$$

Note that $\lambda_k = \varepsilon_k - \varepsilon_0 \sim logistic(0,\tau,0)$. In addition, the expression for the CDF of $\lambda_k + \delta_k$, a linear combination of a logistic and normal random variables, $F_{\lambda+\delta}(z)$, is found in Gupta and Nadarajah (2008). The above probability is rewritten as,

$$Pr_0 = Pr(\lambda_k + \delta_k < V_0 - V_k) = F_{\lambda+\delta}(V_0 - V_k).$$

4 Stopping rule

Note that the observation of search order allows us to estimate the consumer utility in Section 2. Given consumer taste of β_i, the stopping rule allows us to estimate consumer search cost. In addition, it provides more conditions for the outside utility. First, given that consumers kept searching until K, the following must be true,

$$u_0 < z_k,\ k = 1,\cdots,K.$$

Second, given that consumer stopped searching at K, the following must also be true,

$$u_0 > z_m,\ \forall m = K+1,\cdots,J.$$

We find the expression for the probability of $Pr(u_0 < z_k,\ u_0 > z_m)$. To that end, we first compute k's reservation utility,

$$z_k = -c + \int_{z_k}^{\infty} u_k \cdot f(u_k) \cdot du_k,$$

where c is the search cost common to all options. Since the error term of δ_k follows a normal distribution, z_k's reservation utility conditional on ε_k, $z_k|\varepsilon_k$, is expressed as,

$$z_k|\varepsilon_k = V_k + \varepsilon_k + \zeta(\frac{c}{\sigma}) \cdot \sigma. \tag{1}$$

Therefore, z_k's unconditional reservation utility is,

$$z_k = \int_{-\infty}^{\infty} (z_k|\varepsilon_k) \cdot f(\varepsilon_k) \cdot d\varepsilon_k.$$

Note that ε_k is an extreme value type 1 random variable. Evaluation of the above integration leads to

$$z_k = \int_{-\infty}^{\infty} (V_k + \varepsilon_k + \zeta(\frac{c}{\sigma}) \cdot \sigma) \cdot f(\varepsilon_k) \cdot d\varepsilon_k = \int_{-\infty}^{\infty} (V_k + \zeta(\frac{c}{\sigma}) \cdot \sigma) \cdot f(\varepsilon_k) \cdot d\varepsilon_k + \int_{-\infty}^{\infty} \varepsilon_k \cdot f(\varepsilon_k) \cdot d\varepsilon_k.$$

Since the second term in the integration equals to $\gamma \cdot \tau$,

$$z_k = V_k + \zeta(\frac{c}{\sigma}) \cdot \sigma + \gamma \cdot \tau,$$

where γ is an Euler constant. Note that the only unknown parameter in z_k is c. The probability of $\Pr(u_0 < z_k)$ is

$$\Pr(u_0 < z_k) = \Pr(\delta_0 < z_k - V_0) = \Phi(z_k - V_0),$$

where $\Phi$is the normal CDF. Therefore, the overall likelihood is,

$$Pr_i = \prod_{k=1}^{K} \Pr(u_0 < z_k) \cdot \prod_{m=K+1}^{J} \Pr(u_0 > z_m)$$

$$= \prod_{k=1}^{K} \Phi(z_k - V_0) \cdot \prod_{m=K+1}^{J} (1 - \Phi(z_m - V_0)).$$

5 Simulation

Once we estimate the demand primitives of consumer preference and search cost, and utility of outside options, we can conduct simulation studies to estimate incremental sales if the manufacturers changes marketing mix variables such as product features or price. For instance, if some manufacturers lowers their prices, consumers who initially chose outside option may choose to buy, leading to a market expansion. We acheive this using browsing data alone.