Consumer Search and Latent Demand

December 4 2107

Abstract

A typical online consumer just browses and does not make any purchase. In this paper, we estimate the latent demand from such consumer groups through the lens of optimal search theory. In doing so, we demonstrate that consumer browsing histories even without choice are a valuable source of information for firms to understand consumers. As an empirical application, we study the market expansion, 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 both consumers and analysts prior to search but will be known to consumers post search (and during choice). Analysts will never observe this value. We assume that ε_{ij} follows type 1 Extreme Value 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. Consumer i's outside utility is defined as,

$$u_{i0} \sim V_0 + \varepsilon_{i0}.$$

In what follows, we suppress subscript i for clarity.

2 Search Order

Same as in Bart's note. The key point is that the likelihood for the search order can be expressed in terms of β_i and can be estimated accordingly.

3 Choice for outside option

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

$$u_0 > u_j,\ j \in S. \tag{1}$$

Note that the above set of events offers the lower bound for the outside option utility. The corresponding probability is,

$$Pr_0 = Pr(V_j + \varepsilon_j + \delta_j < V_0 + \varepsilon_0).$$

Note that $\lambda_j = \varepsilon_j - \varepsilon_0 \sim logistic(0, \tau, 0)$. A convinient form of CDF for a linear combination of logistic and normal random variable, $F_{\lambda+\delta}(z)$, is found in Gupta and Nadarajah (2008). Therefore, its probability is rewritten as,

$$Pr_0 = Pr(\lambda_j + \delta_j < V_0 - V_j) = F_{\lambda+\delta}(V_0 - V_j).$$

4 Stopping rule

In this section, we show that, given consumer taste of β_i estimated from models of search order, modeling consumers' stopping decisions allow us to estimate search cost. In addition, they offer additional constraints for the estimation of outside uility. Given that consumer continued searching up to K, it must be true that,

$$u_0 < z_k,\ \forall k = 1, \cdots, K. \tag{2}$$

At the same time since consumer stopped searching at K, the following must be also true,

$$u_0 > z_m,\ \forall m = K+1, \cdots, J. \tag{3}$$

Note that the Equations (1), (2), and (3) provide the lower and upper bounds for outside utility given β and c. Our goal in this sub section is to find the corresponding probability expression of $Pr(u_0 < z_k,\ u_0 > z_m)$ that we can use in the estimation. 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 second 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,

$$z_k|\varepsilon_k = V_k + \varepsilon_k + \zeta(\frac{c}{\sigma}) \cdot (\sigma). \tag{4}$$

Accordingly, 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

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

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

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

where γ is an Euler constant. Therefore, we have a closed form solution even if z_k originally involves two dimensional integration. The probability of $\Pr(u_0 < z_k)$ is then,

$$\Pr(u_0 < z_k) = \Pr(\varepsilon_0 < z_k - V_0) = \exp\left(-\exp(\frac{z_k - V_0}{\tau})\right),$$

since ε_0 follows a EV type 1 distribution. Therefore, the overall likelihood for the observed continuation and stopping behavior is,

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

$$= \prod_{k=1}^{K} \exp\left(-\exp(\frac{z_k - V_0}{\tau})\right) \cdot \prod_{m=K+1}^{J} \left(1 - \exp\left(-\exp(\frac{z_m - V_0}{\tau})\right)\right).$$

5 Application

Upon completion of the model estimation, we study the market expansion by changes in products or in price. Such exercises will offer insights to firms on the effect of their product or price changes on sales.