An information search approach to discrete choice analysis

Erlend Dancke Sandorf\textsuperscript{a} Danny Campbell\textsuperscript{a}

\textsuperscript{a}Economics Division, Stirling Management School, University of Stirling

International Choice Modelling Conference, 19–21 August 2019, Kobe
The search for information - the search for alternatives
The role that information and search plays in choice has a long history in economics

- Consideration set models – Richardson (1982), Roberts and Lattin (1991)
- Information acquisition models – Hausmann and Lage (2008), Chorus et al. (2013)
The role that information and search plays in choice has a long history in economics

- Consideration set models – Richardson (1982), Roberts and Lattin (1991)
- Information acquisition models – Hausmann and Lage (2008), Chorus et al. (2013)

People search as long as the expected gain from search exceeds the marginal cost
In many (if not most) choice situations, options are evaluated sequentially.
In many (if not most) choice situations, options are evaluated sequentially.
This means that consideration sets grow sequentially with each period of search.
Econometric model
Utility can be described by a separable and additive utility function:

\[ u_{nis} = \beta \times_{nis} + \varepsilon_{nis} \]

- \( U_{nis} \): Utility
- \( \beta \): Vector of parameters to be estimated
- \( X_{nis} \): Levels of the attributes
- \( \varepsilon \): Type I Extreme value distributed error term with variance \( \pi^2 / 6 \)
The possible gain from search is the difference between any alternative and the current best.

\[ g = u - u_{\text{max}} \]
The value of all possible gains is the area under the “gain” curve

\[ \bar{g} = \int_{-\infty}^{+\infty} gP(g)\,dg, \]
With recall you cannot lose utility by searching for another alternative

\[
g = \begin{cases} 
  u - u_{\text{max}} & \text{if } u \geq u_{\text{max}} \\
  0 & \text{if } u \leq u_{\text{max}} 
\end{cases}
\]
The gain from searching is the area under the “gain” curve above the current best

\[ \bar{g} = \int_{u_{\text{max}}}^{+\infty} (u - u_{\text{max}}) \phi(u) du \]

\[ = \int_{u_{\text{max}}}^{+\infty} u \phi(u) du - \int_{u_{\text{max}}}^{+\infty} u_{\text{max}} \phi(u) du \]

\[ = \phi(u_{\text{max}}) - u_{\text{max}} \int_{u_{\text{max}}}^{+\infty} \phi(u) du \]

where

\[ u_{\text{max}} = (U_{\text{max}} - \mu_t) / \sigma_t \]
An individual will search as long as expected gains are higher than the marginal cost of searching.

\[ \bar{G} - \bar{c} > 0 \]

where

\[ \bar{G} = \bar{g}\sigma \]

i.e. the non-standardized gain to be compared with the marginal cost of search \( \bar{c} \), e.g. time, money, cognitive cost of maintaining a consideration set.
The log likelihood function

\[ P(i_s | C_{ns}) = \prod_{t=1}^{T=J} \left[ \frac{\exp(\beta x_{nis})}{\sum_{j \in C_{ns}} \exp(\beta x_{njs})} \right]^{l_t} \]

where

\[ l_t = \begin{cases} 
1 & \text{if } \bar{G}_t - \bar{c}_t < 0 \quad t = t^* \\
0 & \text{if } \bar{G}_t - \bar{c}_t \geq 0 \quad \forall \ t \neq t^*
\end{cases} \]

and \( t^* \) is the first time the condition is TRUE.
A few challenges to be addressed

- Often the indicator $I_t$ is not observed.
- The parameters $\beta$ enter both the search model and the choice model.
- The sequential nature of the search means the probability of the chosen alternative will always be higher with less search and the conditional probability of search is always decreasing.
An IAL approach

\[ \Pr(C_t) = \int \frac{\exp(\alpha_t)}{\sum_{t=1}^{T} \exp(\alpha_t)} \, d\alpha_t, \alpha_t \sim N(\beta, \sigma) \]

and the joint log-likelihood

\[ \Pr(i) = \sum_{t=1}^{T} \Pr(C_t) \Pr(i, C_t) \]

Evaluating the log-likelihood means evaluating a \( T \)-dimensional integral. Note that this is at the observation level.
Simulating the indicator

1. Take $n$ draws per choice observation from the type I Extreme value distribution
2. For each draw calculate $I_t$
3. Use the average shares of $I_t$ as observation specific weights in the log-likelihood function
Monte-Carlo simulations
Assumptions

- 2000 individuals making 1 choice
- 3, 6, 9 alternatives
- Parameter values and attributes
  - Attribute 1 - 0.4 - (0, 1)
  - Attribute 2 - 0.6 - (0, 1)
  - Attribute 3 - 0.1 - (1, 2, 3, 4)
  - Attribute 4 - -0.7 - (0, 0.2, 0.4, 0.6, 0.8, 1)
- Opportunity cost of time - 0, 0.1, 0.2 - (0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4)
Monte-Carlo simulations - IAL approximation
Failing to consider the sequential nature of the data leads to bias towards zero - a positive parameter.
Failing to consider the sequential nature of the data leads to bias towards zero - a negative parameter.
We tend to under-predict earlier consideration sets and over-predict later ones.
Monte-Carlo simulations - simulating the indicator
Failing to consider the sequential nature of the data leads to bias towards zero.
Failing to consider the sequential nature of the data leads to bias towards zero.
The marginal cost of search appears to be identified with small bias.
It appears that the consideration sets are predicted fairly well.
Concluding remarks
Concluding remarks

- Failing to consider search may lead to underestimation of the choice probabilities
- Failing to consider search may lead to biased estimates
- The IAL model is a good approximation when we cannot observe when people stop search
- Simulating the indicator appears to work well and is a more parsimonious approach compared to the IAL
www.acrg.site

Economics Division
Stirling Management School
University of Stirling
Scotland