Estimating the preference for safety features in cars

Tool: Multinomial Logit and Mixed Logit

The Analytics Edge: Companies such as General Motors carry out conjoint studies with customers to understand the tradeoff (valuation) of different attributes that make up a product or a service. Using data or preferences for safety features in new vehicles and then building models of discrete choice, the company can obtain estimates on the valuation of attributes. This provides important information that can be used to infer the effect of introducing new products in the new market, pricing the product and designing the product features.

The models incorporate the comparison of attributes across alternatives and can also be used to capture heterogeneity in the choice modelling process.

Mixed Logit Model

To capture additional features that multinomial logit cannot capture such as random taste variation.

Standard Logit: $\tilde{U}_{ik} = \beta' x_{ik} + \tilde{\epsilon}_{ik}$

In mixed logit, $\tilde{\beta}$ is modeled as a random parameter, $\tilde{U}_{ik} = \tilde{\beta}' x_{ik} + \tilde{\epsilon}_{ik}$

Standard Logit

$$P(y_i = k) = \frac{e^{\beta' x_{ik}}}{\sum\limits_{l=1}^{K} e^{\beta' x_{il}}}$$

Mixed Logit

$$P(y_i = k) = \int \frac{e^{\beta' x_{ik}}}{\sum\limits_{l=1}^{K} e^{\beta' x_{il}}} f(\beta) d\beta$$

where $f(\beta)$ is the density function of $\tilde{\beta}$. This leads to the integral of logit probabilities.

Mixed logit is computationally more challenging to solve due to the use of simulation optimization methods.

The problems are no longer convex in this setting and finding a global optimum might not be easy. From an estimation perspective, the goal is to find the parameters θ that define the density function $f(\beta|\theta)$ where the functional form $f(.)$ is given but parameters θ are unknown. To approximate the probability value given a particular θ, draw multiple β_r vectors from the distribution $f(\beta|\theta)$.

$$P(y_i = k) \approx \sum_{r=1}^{R} \frac{e^{\beta_r' x_{ik}}}{\sum\limits_{l=1}^{K} e^{\beta_r' x_{il}}}$$

Plugging this approximation into the log-likelihood objective function and estimate θ value by doing optimization.

For mixed logit with repeated choices (panel data), where i: individual, k: alternative, t: observation

$$P(y_{i1} = k_1, \ldots, y_{iT} = k_T) = \int \prod_{t=1}^{T} \left(\frac{e^{\beta' x_{ik_i t}}}{\sum_{l=1}^{K} e^{\beta' x_{ilt}}} \right) f(\beta) d\beta$$

Panel data needs to account for the fact that the errors are correlated for the same individual over time.

Which of the following packages would you prefer the most?

Choose by clicking one of the buttons below

Full Speed Range Adaptive Cruise Control	Adaptive Cruise Control	Full Speed Range Adaptive Cruise Control	NONE: I wouldn't purchase any of these packages
-	Navigation System with Curve Notification & Speed Advisor	Traditional Navigation System	
Traditional Back Up Aid	Rear Vision System	-	
Lane Departure Warning	-	-	
Front Collision Warning	-	Front Collision Warning	
Side Head Air Bags	Side Body Air Bags	Side Body & Head Air Bags	
Emergency Notification with Pictures	Emergency Notification	-	
-	-	Night Vision with Pedestrian Detection	
-	Head Up Display	Head Up Display	
Option Package Price: $3, 000	Option Package Price: $500	Option Package Price: $12, 000	
○	○	○	○