РАЗДЕЛ I

Пазари, управление и иновации в икономиката на знанието

МОДЕЛИРАНЕ НА ВЕРОЯТНОСТНИЯ ИЗБОР ПРИ ПОКУПКА

Доц. д-р Тодор Кръстевич

Резюме

В това изследване се систематизират теоретичните основи на математическото моделиране на потребителското поведение в процеса на избор в контекста на теорията на случайната полезност. Фокусът е насочен върху някои аспекти на заявения дискретен избор, които се поддават на иконометрично оценяване. Основната авторова теза е, че, изхождайки от теорията на случайната полезност, е възможно да се създаде инструментариум за прогностичен анализ на търсенето. Целта на изследването е разработване на прозрачна и проследима аналитична методология за анализ на емпирични данни от заявени предпочитания, осигуряваща възможност за оценяване, анализиране и предсказване на потребителския избор. За реализирането на тази цел е направен кратък синопсис на теорията на дискретния избор, както и обзор на методите за оценка и анализ на традиционните иконометрични модели с дискретни зависими променливи. Особено внимание се обръща на проблемите при специфицирането на адекватни дихотомни и мултиномиални логит модели, които са сърцевината на аналитичния инструментарий и чието оценяване и валидиране е постижимо с помощта на софтуер с отворен код. Чрез систематизиране на постиженията в областта се извеждат насоки за използване на моделите за целите на прогностичния анализ, свързан с решаването на маркетингови проблеми. Предлагат се техники за прогностичен анализ на потребителския избор с помощта на изборни модели, в т.ч. извеждане на агрегирани прогнозни оценки, съставяне на пазарни симулатори за предиктивно моделиране на пазарните дялове, оценяване на готовността за плащане и ценовата еластичност на ниво изборни алтернативи.

Ключови думи*:* потребителски избор, модели с дискретния избор, теория на случайната полезност, дихотомен и мултиномиален логит, прогностичен анализ на потребителското търсене.

JEL: C25, C52, M31, D46.

MODELING PROBABILISTIC CHOICES WHEN BUYING

Assoc. Prof. Dr. Todor Krastevich

Abstract

This study systematizes the theoretical foundations of mathematical modeling of consumer behavior in the choice process in the context of random utility theory. The main attention is paid to some aspects of econometric estimation of discrete choice models. The main thesis is that based on the theory of random utility, you can create tools for predictive analytics of demand. The aim of the study is to develop a transparent and traceable analytical methodology for analyzing empirical data from stated preferences, providing opportunities for evaluating, analyzing and predicting consumer choice. To achieve this goal, a brief synopsis of the discrete choice theory was made, as well as a review of methods for evaluating and analyzing traditional econometric models with discrete dependent variables. Particular attention is drawn to the problems in specifying adequate binary and multinomial regression models, which are the core of analytical tools, and whose evaluation and verification is feasible with the help of open source software. Systematizing achievements in the field, recommendations are given for using predictive models for solving marketing problems. Methods for predictive analytics of consumer choice are proposed, including output of aggregated forecast estimates, building of market simulators for predictive modeling of market share, assessment of willingness-to-pay and price elasticity at the level of product alternatives.

Keywords: consumer choice behavior, stated choice models, random utility, binary and multinomial logit, predictive analytics of customer demand.

JEL: C25, C52, M31, D46.

Увод

Разбирането и измерването на ефектите от потребителския избор е един от най-интересните и същевременно сложни за разбиране аспекти на маркетинговите изследвания. Изборът се проявява в много разновидности и форми. Може да бъде дискретен, в смисъл на избор само на един елемент, или може да бъде непрекъсната, когато се купуват или избират няколко елемента. Изборът може да отразява мисленето, навика или спонтанната реакция на потребителя към маркетинговите променливи. Той невинаги води до покупки и невинаги се подчинява на рационалните презумпции за полезност. Изборът обикновено е резултат от компромиси, които в същността си могат бъдат както компенсаторни, така и некомпенсаторни. Изучаването и предсказването на потребителския избор е едно от най-важните предизвикателства за маркетинга, тъй като се поддава на въздействие и контрол чрез маркетинговите инструменти (Chandukala et al., 2007, p. 99). Това може да се обоснове с обстоятелството, че съвкупността от изборните решения за покупка формират агрегираното търсене. Точните прогнози за търсенето и пазарните дялове са от решаващо значение както за маркетинговата ефективност на бизнес организациите, така и за управлението и разпределението на публични ресурси. Конкретни приложения на подобни прогнози могат да се търсят при прогнозиране търсенето на нов продукт при алтернативни ценови стратегии, при изготвянето на бизнес планове за въвеждането на нови технологии, при анализ на въздействието на сливането или придобиването на компании върху пазарните дялове, при прогнозиране на трафика и пътникопотоците в транспортни системи при алтернативни форми на организация, транспортни услуги и средства, както и при и анализа на конкурентни сценарии за въвеждане на нови телекомуникационни услуги.

Основната теза на настоящото теоретично изследване е, че, изхождайки от теорията на случайната полезност, е възможно да се създаде инструментариум за прогностичен анализ на търсенето. Целта е разработване на прозрачна и проследима аналитична методология за анализ на емпирични данни от заявени предпочитания, осигуряваща възможност за оценяване, анализиране и предсказване на потребителския избор. За изпълнението на тази цел адаптираме методологически инструментариум за анализ на дискретния избор. Чрез моделиране на изборното поведение на потребителите се опитваме да изведем агрегирани модели на пазарната структура и да прогнозираме маркетинговото въздействие върху търсенето при различни сценарии.

Моделите с дискретен избор са иконометрични модели, които описват, обясняват и предсказват избор между две или повече алтернативи (т.е. когато наборът от алтернативи не е повече от изброим). Моделите на дискретен избор позволяват, въз основа на някои характеристики (атрибути) на даден икономически субект или ситуация да се оцени вероятността за избор на една или друга алтернатива. В тази студия се разглеждат основните теоретични принципи на дискретния избор, както и някои базисни математически модели и алгоритми, необходими за прилагане на това знание на базата на заявени предпочитания, изведени от емпирично наблюдаваеми изборни решения от страна на респонденти, направени при експериментални условия.

1. Теоретични основи на дискретния избор

Моделирането на човешкото поведение в процеса на избор се основава на постижения в областта на психологията и поведенческата икономика. Теоретичните корени на моделирането на дискретния избор се откриват в (1) математическата психология и по-конкретно – в теорията на случайната полезност, свързвана с имената на Р. Дънкан Лус (Luce, 2014) и Антони Марли (Marley, 1968), както и в (2) микроикономическите теории на потреблението, фокусирани върху изследване и обясняване поведението на икономическите субекти (индивиди, домакинства, организации) при вземане на решения чрез дезагрегирани модели на търсенето (Bierlaire & Lurkin, 2017).

Традиционната микроикономическа парадигма за изследване на потребителския избор е детерминистичната. Според тази парадигма индивидите се държат рационално и детерминирано. Предполага се, че вземащите решения имат перфектна способност да разграничават продуктовите алтернативи, могат да обработват информация, да избират най-добрия вариант за покупка и да повтарят този идентичен избор при повтарящи се идентични обстоятелства. Това поведение се подразбира от аксиоматичните постулати, които характеризират допускането за „рационалност“, а именно – пълнота, транзитивност и непрекъснатост (Nicholson & Snyder, 2017, p. 89).

За формалното описание на постулатите за рационалност на потребителското поведение ще си послужим с набор от абстрактни математически обозначения. Нека отбелязваме с X множеството от алтернативни варианти на продукти, марки или действия, измежду които хипотетични потребители са в състояние да избират. Целта е да определим кой от елементите на множеството X ще бъде избран/закупен от даден потребител. Допускаме, че всеки конкретен потребител би избрал този вариант от множеството X, към който има най-силни предпочитания. За обозначаване на посоката на предпочитание и/или безразличие между две изборни алтернативи ще използваме операторите $\succcurlyeq$, $\preccurlyeq$, $\succ$, $\prec$ и $\sim$. С тези оператори изразяваме посоката на предпочитанието, но не и неговата интензивност (т.е. изразяваме невъзможността да се определи емпирично, с колко по-силно се предпочита една, спрямо друга/други алтернативи.

Нека допуснем, че две продуктови алтернативи, a и b, принадлежат към множеството X (т.е. $a, b \in X$). С израза $a \succcurlyeq b$ бихме могли да отбелязваме ситуация, при която потребителят предпочита a пред b, или има абсолютно еднакви предпочитания (индиферентност) към двете алтернативи. Ако потребителят е единствено и само индиферентен между двете алтернативи, то бихме могли да използваме индикатора $a \sim b$, което ще бъде равносилно на $a \succcurlyeq b \wedge a \preccurlyeq b$. И обратното, ако потребителят стриктно предпочита алтернатива b пред алтернатива a, то следва да използваме оператора $a \prec b$.

Ако изходим от допускането, че всеки потребител се държи рационално, то би следвало, неговите предпочитания да съответстват на допусканията за пълнота и транзитивност в рамките на множеството от алтернативи X. Допускането за *пълнота* означава, че за всяко $a, b \in X$ е възможно да се установи дали $a \succcurlyeq b$ е вярно или не е вярно. Допускането за *транзитивност* означава, че за всяко $a, b, c \in X$, ако $a \succcurlyeq b$ и $b \succcurlyeq c$, то $a \succcurlyeq c$. Допускането за *непрекъснатост* означава, че ако даден потребител докладва, че предпочита алтернатива a пред алтернатива b, то и всички други алтернативи, „близки“ до a, би трябвало да бъдат предпочитани пред b.

Оттук, ако с $u(.)$ отбелязваме възприеманата от потребителя полезност на конкретна продуктова, принадлежаща към множеството X, то структурата на предпочитанията на потребителя би могла да се представи чрез *функция на полезността*, т.е., ако:

$$a \succcurlyeq b \Leftrightarrow u(a) \geq u(b).$$

Пълнотата и транзитивността на предпочитанията гарантират съществуването на функцията на полезността. Презумпцията за непрекъснатост на функцията на полезността пък от своя страна допуска и диференцируемост.

Като се имат предвид предположенията за пълнота, транзитивност и непрекъснатост на функцията на полезността, възможно е формално да се покаже, че потребителите са в състояние да класифицират всички възможни ситуации от най-малко желаните до най-желаните. По-желаните ситуации предлагат повече полезност, отколкото по-малко желаните (Nicholson & Snyder, 2017, p. 90).

Оттук, ако приемем, че всеки потребител се стреми да „максимализира“ полезността, то процесът на вземане на решение за покупка би следвало да се представи като решение на следната целева функция:

$$u(q) \rightarrow max! \quad \text{... като } q \in X.$$

Ако всяка една изборна алтернатива q от множеството X е независима от другата и може да бъде избирана/закупена като цяло, а не на части и/или в различни обеми, то говорим за ситуация на *дискретен* избор.

2. Модели на дискретния избор

Моделите на дискретния избор са икономически (иконометрични) модели, които описват, обясняват и предсказват избор между две или повече алтернативи и то, когато наборът от изборни алтернативи е краен и изброим. Моделите на дискретен избор позволяват въз основа на някои релевантни характеристики (атрибути) на даден икономически субект или ситуация да се оцени вероятността за избор на една или друга алтернатива.

Изследването на поведението при дискретен избор се описва от (1) обектите на избор и наборите от алтернативи, достъпни за вземащите решения, (2) наблюдаваните атрибути на вземащите решения и (3) модела на индивидуален избор и поведение и разпределението на поведенческите модели сред изучаваната популация от потребители. Предполага се, че наблюдаваните емпирични данни се генерират чрез случаен подбор на индивида от целевата съвкупност, съдържат потребителските му характеристики, набора от достъпни продуктови алтернативи за всеки отделен потребител и действителния му избор. Извадката се получава в резултат на последователност от независими тестове, със или без повторения, при които се наблюдава последователност от избори за индивиди със същите измерими атрибути и алтернативни изборни множества (D. L. McFadden, 1974, pp. 106–107).

За да дефинираме операционално модела на дискретния избор, приемаме следните обозначения и понятия (Bierlaire & Lurkin, 2017, pp. 52–53):

n: потребител, вземащ решение за покупка.

i: продуктова алтернатива, обект на изборно решение ($i = 1, \dots, J$).

y_j: вектор с наблюдавани изборни решения $(y_1, \dots, y_J)$, $y_j \in \{0,1\}$, $\sum_j y_j = 1$.

C_n: изборно множество, представляващо краен брой дискретни възможности за избор, вземани под внимание от потребител n в процеса на вземане на решение за покупка.

J_n: брой алтернативи (т.е. възможности за избиране) в C_n (напр. начините на транспортиране в градска среда, достъпни за потребител n при ежедневното му придвижване до работното му място).

U_{in}: полезност, която потребител n асоциира с изборна алтернатива i (напр. възприеманата от потребител n полезност от използване на конкретен вид градски транспорт за придвижване до работното място).

z_{in}: вектор с атрибутивните характеристики/свойства на алтернатива i, достъпна за потребител n (напр. време за пътуване, цена за едно пътуване, комфорт и сигурност за всеки един начин на придвижване до работното си място). Атрибутивните свойства биха могли да бъдат общи (т.е. отнасящи се до всички възможни продуктови алтернативи в изборните множества), както и специфични (т.е. отнасящи се само до някакво подмножество от продуктови алтернативи или дори до конкретна продуктова алтернатива).

s_n: вектор със социално-икономически, демографски и/или психографски характеристики на потребител n (напр. доходи, възраст, образование, професия, цел на покупката, нагласи, убеждения и др.).

Отделните потребители е възможно да попадат в различни изборни ситуации C_n и биха могли да имат различни вкусове и предпочитания спрямо едни и същи продуктови алтернативи. Например, две домакинства, с различни доходи и различни локации на местоживеене е твърде вероятно да имат различни изборни множества (например, различни начини за придвижване до определена локация в населеното място или различни местни доставчици на фиксирани телекомуникационни услуги), както и да отдават различно относително значение на отделните продуктови атрибути, съставящи изборните алтернативи.

При моделирането на дискретния избор се приема, че всеки потребител n избира само една продуктова i от наличните J_n в изборно множество C_n. Допуска се, че алтернатива $i \in C_n$ ще бъде избрана от потребител n, ако нейната възприемана полезност е по-висока от възприеманата полезност на останалите алтернативи в изборното множество, т.е.:

$$U_{in} \geq U_{jn}, \quad \forall_j \in C_n.$$

Тази презумпция би била валидна, само ако индивидите се държат рационално и детерминирано. Предполага се, че вземащите решения имат перфектна способност за дискриминация между всички налични продуктови алтернативи, могат да обработват перфектно цялата налична информация, винаги да избират най-добрия вариант за покупка и стриктно да повтарят този избор при идентични обстоятелства. Това се подразбира и от предполагаемите свойства на предпочитанията – пълнота, транзитивност и непрекъснатост. Подобни предположения обаче едва ли биха могли да бъдат приети като съответстващи на реалното, наблюдаваемо поведение на потребителите. По тази причина се появява и налага теорията за вероятностния избор (D. McFadden, 1980). Основното теоретично допускане е, че процесът на избор се състои от два отделни независими помежду си подпроцеса – детерминистичен (систематичен) и стохастичен (случаен). Признава се, че резултатът от изборния процес може да се предскаже само чрез определяне на вероятността на различните изборни резултати, отчитайки конкретната изборна ситуация и спецификата на вземащия решение. За да се операционализира тази теоретична концепция се приема, че източникът на стохастичността се дължи на грешки, направени от анализиращия при разработването на модела. Предполага се, че докато потребителите са детерминирани и рационални „максимизатори“ на полезността, анализаторите не са в състояние напълно да разберат и моделират всички релевантни фактори, влияещи върху човешкото поведение. Дори да допуснем, че индивидът е „рационален“, всезнаещ и винаги избира алтернативите с най-голямата полезност, тези полезности не са известни със сигурност на анализиращия и следва да се разглеждат като случайни променливи. Именно поради тази причина този подход за моделиране на потребителския избор се нарича теория на случайната полезност. Основното научно „достойнство“ на теорията на случайната полезност е, че тя е теоретично обоснована от гледна точка на микроикономиката и позволява връзки с концепции и техники за разработване и използване на статистически модели.

В този контекст полезността U_{in}, която потребител n асоциира с изборна алтернатива i е на практика неизвестна за изследователя и следва да бъде обект на емпирично оценяване или предсказване. За целта се приема условно, че U_{in} е непрекъсната *случайна* ненаблюдаваема променлива, дефинирана като:

$$U_{in} = V_{in}(z_{in}, s_n) + \varepsilon_{in}.$$

Моделът се състои от два основни компонента. Компонентът $V(z_{in}, s_n)$ представлява детерминистична (наричана още и систематична) част, която е по същество функция от продуктовите атрибути и характеристиките на потребителите, поддаващи се на пряко емпирично наблюдение. Другият компонент е стохастичен, изразява случайните грешки от измерването и представлява съвместно статистическо разпределение на случайния вектор $\varepsilon_{in} = (\varepsilon_{1n}, \dots, \varepsilon_{J_n n})$. Именно поради наличието на стохастичен компонент всеки изборен модел представлява вероятностен модел и се основава на теорията на случайната полезност (вж. Koppen, 2001, pp. 1646–1651).

Изхождайки от по-горните разсъждения, вероятността P_n, с която даден потребител n би избрал алтернатива i от известно му изборно множество C_n, може да се изрази формално като[1]:

$$P_n(i|C_n) = \Pr(U_{in} \geq U_{jn}, \forall_j = 1, \dots, J_n \in C_n),$$

като сумата от изборните вероятности за всяка продуктова алтернатива $i \in C_n$ е винаги равна на единица (т.е. е в сила $\sum_{i \in C_n} P_n(i|C_n) = 1$). И тъй като стохастичният компонент ε_{in} е непрекъсната случайна променлива, то вероятността за възникване на две продуктови алтернативи с една и съща изборна вероятност е на практика нулева.

От изложените постановки от теорията на случайната полезност може да се изведе обобщението, че в процесите на потребителския избор се разграничават следните основни елементи поддаващи се на емпирично наблюдение:

- Изборни алтернативи i (напр. продукти), описвани чрез комбинация от продуктови атрибути.
- Изборни множества C_n, съставени от продуктови алтернативи.
- Потребители n, които вземат решения и избират продуктова алтернатива от известните им и достъпни изборни множества.
- Демографски, социално-икономически и/или психографски характеристики на потребителите s_n.
- Правило за вземане на решение за избор, което следва отделния потребител.

В съответствие с теорията на случайната полезност всеки потребител се стреми към максимализиране на полезността, т.е. следва правилото „избирам алтернативата с максималната“ полезност, което се поддава сравнително лесно на статистическо описание и емпирично изучаване. Затова фокусът на следващото изложение ще бъде насочен именно към адекватно моделиране на „правилата“ за изборните решения с помощта на

[1] За подробно математическо извеждане на общия модел на случайната полезност вж. Bierlaire & Lurkin, (2020).

емпирични иконометрични модели с цел адекватно описание и надеждно предсказване на потребителския избор.

2.1. Моделиране на избора при две алтернативи

В тази подточка ще се опитаме да изясним как се идентифицира и оценява модел на дискретен избор, включващ само две взаимоизключващи се изборни алтернативи. Подобни модели се наричат двоични, дихотомни или бинарни. В изложението ще използваме тези определения като взаимнозаменяеми. Особено внимание ще отделим на изясняването на структурата на дихотомните моделите, съставена от детерминистичен и стохастичен компонент, както и на числовите алгоритми, допустими при калибрирането на параметрите на моделите въз основа на емпирични данни.

2.1.1. Общ вид на дихотомния модел

Според теорията на случайната полезност всеки индивид–потребител n в процеса на вземане на решение за покупка попада в изборна ситуация C_n, състоящи се от краен брой дискретни продуктови алтернативи J_n, от които той избира тази алтернатива i, чиято полезност U_{in}^{max} се възприема от него като най-висока. Поради вече изложените съображения полезността се разглежда като случайна ненаблюдаваема променлива, състояща се от детерминистичен и стохастичен компоненти. Допускаме, че избирането на максималната полезност е имплицитната цел на потребителя във всяка една изборна ситуация.

Като частен случай на изборните ситуации може да се разглежда сценарий, при който изборното множество се състои само от две алтернативи (напр. купувам или не купувам, или придвижване до работното място с личен автомобил или с градски транспорт). Логиката на съставяне на дихотомен изборен модел, съответстващ на този сценарий, е следната. Нека опишем едно абстрактно изборно множество, състоящо се от две продуктови алтернативи, i и j, с обозначението:

$$C_n = \{i, j\}.$$

Приемаме, че всяка от двете изборни алтернативи има ненаблюдаваема полезност U_{in} и U_{jn}, които съдържат както детерминистичен (систематичен) компонент V_{in}, респ. V_{jn}, така и стохастичен компонент ε_{in}, респ. ε_{jn}:

$$U_{in} = V_{in} + \varepsilon_{in}, \text{ респ. } U_{jn} = V_{jn} + \varepsilon_{jn}.$$

Припомняме, че детерминистичните компоненти V съдържат информацията, която е възможно да бъде емпирично наблюдавана от изследователя, докато стохастичните компоненти изразяват информация, която не може да бъде наблюдавана емпирично от анализиращия, както и грешки при измерването и/или идентификацията на самия модел. Оттук моделът на избора следва да представлява вероятността P, с която даден потребител n би избрал алтернатива i. Това на практика означава да се оцени вероятността $\Pr(.)$, с която възприеманата от потребител n полезност на алтернатива i е по-висока от възприеманата полезност на алтернатива j:

$$P_n(i|\{i,j\}) = \Pr\left(U_{in} \geq U_{jn}\right).$$

Ако заместим изразите на полезност на двете алтернативи в последното уравнение, ще получим:

$$\begin{aligned} P_n(i|\{i,j\}) &= \Pr\left(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{in}\right) \\ &= \Pr\left(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}\right) \\ &= \Pr\left(\varepsilon_n \leq V_{in} - V_{jn}\right), \quad \text{като } \varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}. \end{aligned}$$

Последният израз на практика показва, че вероятността да бъде избрана продуктовата алтернатива i от потребител n, е равна на вероятността, ε_n да е по-малка от или равна на разликата между детерминистичните компоненти на алтернатива i и алтернатива j. Това заключение е изключително важно за предсказване на изборната вероятност, тъй като детерминистичните компоненти се поддават на пряко емпирично наблюдение.

По-сложно е оценяването на стохастичните компоненти ε_{in} и ε_{jn} на модела. Тъй като тези случайни променливи не се поддават на емпирично наблюдение, остава алтернативата да приемем някаква хипотеза за техните статистически разпределения и по-конкретно – какво е математическото очакване (т.е. очакваната средна стойност), каква е дисперсията и какъв е видът на статистическото им разпределение. Обръщаме внимание, че при дихотомна ситуация е необходимо да се приеме хипотеза за математическото очакване само на разликата ε_n. При ситуация с повече от две изборни алтернативи се налага да се правят допускания за всяка отделна продуктово специфична стохастична променлива ε_j.

Математическо очакване на стохастичните компоненти

Нека допуснем, че очакваните средни стойности на двете стохастични променливи ε_{in} и ε_{jn} са съответно равни на β_{in} и β_{in}, т.е.

$$E[\varepsilon_{in}] = \beta_{in} \ \text{ и } \ E\left[\varepsilon_{jn}\right] = \beta_{jn}.$$

Нека дефинираме и две помощни променливи, ε'_{in} и ε'_{jn}, за които да е в сила:

$$\varepsilon'_{in} = \varepsilon_{in} - \beta_{in} \text{ и респ. } \varepsilon'_{jn} = \varepsilon_{jn} - \beta_{jn}.$$

От това би следвало, че очакваните стойности на помощните променливи ще са:

$$E[\varepsilon'_{in}] = E\left[\varepsilon'_{jn}\right] = 0.$$

Заместваме тези помощни променливи в първоначалния вид на основния модел и предприемайки някои елементарни преобразувания, които не водят до промяна на изборната вероятност:

$$\begin{aligned} P_n(i|\{i,j\}) &= \Pr\left(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}\right) = \\ &= \Pr\left(V_{in} - V_{jn} \geq \varepsilon'_{jn} + \beta_{jn} - (\varepsilon'_{in} + \beta_{in})\right) = \\ &= \Pr\left(V_{in} + \beta_{in} - \left(V_{jn} + \beta_{jn}\right) \geq \varepsilon'_{jn} - \varepsilon'_{in}\right) = \\ &= \Pr\left(V_{in} + \beta_{in} - \left(V_{jn} + \beta_{jn}\right) \geq \varepsilon'_{n}\right), \end{aligned}$$

където $\varepsilon'_n = \varepsilon'_{jn} - \varepsilon'_{in}$.

От последния израз е видно, че очакваните средни стойности на случайните променливи са прехвърлени в детерминистичната (систематичната) част на уравнението. Това позволява тяхното обхващане с помощта на параметър, който може да бъде оценен на базата на емпирични данни. Прието е, този параметър да се нарича „специфична за алтернативата константа“. Тази константа „улавя“ средния ефект върху полезността на всички фактори, които не са включени в модела и се отнася само за конкретната продуктова алтернатива. По този начин тя изпълнява функция, подобна на свободната константа в регресионния модел, която също отразява средния ефект на всички невключени независими променливи. Когато в изборния модел са включени специфични за алтернативата константи, ненаблюдаваемата част от полезността ε_{jn} (респ. ε_{ij}) има очаквана стойност равна на нула (Train, 2009, p. 20).

Дихотомният вероятностен изборен модел притежава някои интересни свойства. Например той не е чувствителен от равномерно изместване на всички функции на полезността. Това означава, че ако от всяка оценка на полезност на съответната продуктова алтернатива i и j бъде добавена или извадена една и съща константна стойност (напр. K), изборната вероятност не се променя:

$$P_n(i|\{i,j\}) = \Pr\left(U_{in} \geq U_{jn}\right) = \Pr\left(U_{in} + K \geq U_{jn} + K\right) \quad \dots \quad \forall\, K.$$

Това свойство позволява различни еквивалентни преобразувания. Така например, ако в обичайното изразяване на полезността на двете разглеждани продуктови алтернативи i и j преместим очакваните средни стойности β_{in} и β_{jn} на случайните компоненти в детерминистичната част на уравненията (вж. спецификация А), то тази спецификация е напълно еквивалентна както на спецификация Б, така и на спецификация В:

Спецификация А:	$\Leftrightarrow$	Спецификация Б:	$\Leftrightarrow$	Спецификация В:
$U_{in} = V_{in} + \beta_{in} + \varepsilon'_{in}$		$U_{in} = V_{in} + \varepsilon'_{in}$		$U_{in} = V_{in} + \beta_{in} - \beta_{jn} + \varepsilon'_{in}$
$U_{jn} = V_{jn} + \beta_{jn} + \varepsilon'_{jn}$		$U_{jn} = V_{jn} + \beta_{jn} - \beta_{in} + \varepsilon'_{jn}$		$U_{jn} = V_{jn} + \varepsilon'_{jn}$

Вследствие на това свойство при практическото оценяване на изборните модели често се пристъпва към нормализиране (т.е. приравняване) на една от константите на нула и оценяване само на един параметър, β_n, представляващ разликата между $\beta_{in} - \beta_{jn}$.

Друго интересно свойство на изборния модел е независимостта му от мащаба. Това означава, че изборната вероятност не се променя при равномерно мащабиране на всички функции на полезността. С други думи, ако умножим всяка функция на полезността U_{in} и респ. U_{jn} с едно и също положително число (напр. α), то няма изменение в P_n(i|{I,j}). Във формален вид това съответства на следните релации:

$$P_n(i|\{i,j\}) = \Pr\left(U_{in} \geq U_{jn}\right) = \Pr\left(\alpha U_{in} \geq \alpha U_{jn}\right) \qquad \forall\, \alpha > 0.$$

Дисперсия на стохастичните компоненти

Дисперсията определя мащаба на разсейването на ненаблюдаваемите случайни величини и/или грешки в модела. За разлика от математическото очакване за средната, дисперсията на стохастичния компонент не се поддава на идентификация, тъй като варира при отделните наблюдения (Swait & Louviere, 1993, p. 305). Това означава, че хипотетично всяка една произволно избрана дисперсия би могла да бъде допустима и че не съществува начин за идентифициране на дисперсията на случайния компонент на базата на емпирични данни. Затова при практическото оценяване на изборни модели често дисперсията се нормализира към 1, въпреки че това не позволява отчитането на хетерогенността между потребителите (Louviere & Eagle, 2006).

Разпределение на стохастичните компоненти

Поредното допускане е свързано с вида на статистическото разпределение на стохастичните компоненти. Изхожда се от хипотезата, че ε_{in} и ε_{jn} представляват максимума на множество случайни променливи, отразяващи влиянието на ненаблюдаваеми атрибути (напр. опитът или настроението на респондентите), грешки при измерването или грешки, произтичащи от самата спецификация на модела. Тази хипотеза е в съответствие и с теорията за максимализиране на полезността. От статистическата теория е известно, че максимумът на множество от идентични, независимо разпределени случайни променливи, приблизително следва разпределение на екстремните стойности, частен случай на което е разпределението на Гумбел (Alves & Neves, 2011, p. 493; Gumbel, 1954, pp. 19–46) , характеризиращо се с математическо очакване (местоположение) η и мащаб (дисперсия) μ, като $\mu > 0$. Разпределението на екстремните стойности на Гумбел обикновено се изписва като $EV(\eta,\mu)$. Функцията на вероятностната плътност на това разпределение е следната:

$$f(t) = \mu e^{-1(t-\eta)} e^{e^{-mu(t-\eta)}}.$$

Чрез интегриране се извежда и кумулативната функция на вероятностното разпределение, която има вида:

$$P(c \geq \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt = e^{-e^{\eta(c-mu)}}.$$

Видно е, че последната функция има затворена форма. Това е много удобно свойство на кумулативната функция, което улеснява впоследствие решаването на числовия проблем.

Функцията на плътността на вероятностното разпределение на екстремните стойности е положително асиметрично (дясно изтеглено). Това разпределение има някои интересни свойства. Ако приемем, че стохастичният компонент на изборния модел следва разпределението на Гумбел:

$$\varepsilon \sim EV(\eta,\mu),$$

то местоположението (математическото очакване на средната $E[\varepsilon]$) и мащабът (дисперсията $var[\varepsilon]$) ще са съответно:

$$E[\varepsilon] = \eta + \frac{\gamma}{\mu} \quad \text{и} \quad var[\varepsilon] = \frac{\pi^2}{6\mu^2},$$

като γ е константата на Ойлер-Маскерони, която е равна на:

$$\gamma = -\int_0^\infty e^{-x} \ln x \, dx \approx 0{,}5772.$$

Допускайки хипотезата, че значенията на стохастичните компоненти на дихотомния изборен модел ε_{in} и ε_{jn} следват разпределението на Гумбел, за удобство може да допуснем, че параметърът на местоположението (очакваната стойност) е равен на нула. Така разпределенията могат да се опишат като $\varepsilon_{in} \sim EV(0, \mu)$ и респ. $\varepsilon_{jn} \sim EV(0, \mu)$.

Припомняме обаче, че в дихотомния изборен модел присъства компонента ε_n, представляващ разликата $\varepsilon_{jn} - \varepsilon_{in}$. За разлика от нормалното статистическо разпределение, разликата между две разпределения на екстремните стойности на Гумбел не води до ново разпределение на екстремни стойности, а представлява логистично статистическо разпределение (Nadarajah & Kotz, 2005, p. 3169), т.е.:

$$\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in} \sim Logistic(0, \mu).$$

Функцията на плътността на вероятностното разпределение на логистичното статистическо разпределение се изразява като:

$$f(t) = \frac{\mu e^{-\mu(t-\eta)}}{\left(1 + e^{-\mu(t-\eta)}\right)^2},$$

а съответната й кумулативна функция на логистичното разпределение:

$$P(c \geq \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt = \frac{1}{1 + e^{-\mu(c-\eta)}},$$

където $\mu > 0$. По подобие на функцията на разпределението на екстремните стойности и кумулативната функция на логистичното вероятностно разпределение има затворена форма.

Изхождайки от коментираните допускания за детерминистичните и стохастичните компоненти до момента, можем да заключим, че дихотомният изборен модел

$$P_n(i|\{i; j\}) = \Pr\left(\varepsilon_n \leq V_{in} - V_{jn}\right) = F_\varepsilon(V_{in} - V_{jn})$$

изразява вероятността P, с която даден потребител n би избрал продуктова алтернатива i вместо j тя е равна на вероятността ε_n да е по-малка или равна на разликата между детерминистичните компоненти V_{in} и V_{jn}. Тъй като вече знаем, че ε_n следва логистично разпределение и сме наясно с израза на кумулативната функция на разпределение на ε_n, операционалното извеждане на дихотомния изборен модел е тривиално:

$$P_n(i|\{i; j\}) = \frac{1}{1 + e^{\mu(V_{in} - V_{jn})}} = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}}.$$

Последният израз е познат като дихотомен (или двоичен) логит модел и чрез него лесно може да се изчисли (и респ. предскаже) вероятността, с която отделният индивид n би избрал продуктовата алтернатива i. В него параметърът за мащаба μ (изразяващ дисперсията на разпределението на грешките) обикновено се приема за единица (Whitten, 2004, p. 8).

Логит моделът би могъл да бъде оценен с помощта на логистична регресия, при която се приема, че емпирично наблюдаваните изборни решения на потребителите са в съответствие с аксиомата за независимост от ирелевантните алтернативи[2]. Според тази аксиома вероятността за избор на конкретна продуктова алтернатива не бива да се влияе от включването или пропускането на други алтернативни опции в съответното изборно множество. Ако се установи нарушение на тази хипотеза, следва да се използват други варианти на модела (напр. мултиномиален пробит, вложен логит или логит със случайни параметри). В следващото изложение ще бъде разгледан основният числов алгоритъм, приложим при оценяването на дихотомния логит модел.

2.1.2. Оценяване на дихотомния модел

Най-често използваният числов алгоритъм за оценяване параметрите на изборните модели е методът на максималното правдоподобие. За прилагането му са необходими данни от емпирично наблюдавани изборни решения y_j измежду краен брой J_n продуктови алтернативи, формиращи изборни множества C_n от страна на извадка от респонденти N (като $n = 1, \ldots, N$). Освен това са необходими данни z_{in} и z_{jn} за общи или специфични атрибутивни характеристики на всяка от изборните алтернативи, достъпна за съответните потребител n, както и (евентуално) вектори s_n с данни за социално-икономически, демографски и/или психографски характеристики на всеки потребител n от извадката. За удобство тези данни могат да бъдат описани като два вектора (независими променливи в модела на всяка една от двете алтернативи) по следния начин:

$$x_{in} = h(z_{in}, s_n) \qquad \text{и} \qquad x_{jn} = h(z_{jn}, s_n).$$

Допускайки, че параметърът за мащаба μ в дихотомния изборен модел е 1, то общият му вид може да се запише като:

$$P_n(i|\{i;j\}) = \frac{e^{V_{in}}}{e^{V_{in}} + e^{V_{jn}}},$$

където детерминистичните компоненти V_{in} (респ. V_{jn}) се изразят като линейна комбинация от z_{in} (респ. z_{jn}) и s_n:

$$V_{in} = \sum_{k=1}^{K} \beta_k x_{ink} \qquad\qquad V_{jn} = \sum_{k=1}^{K} \beta_k x_{jnk}.$$

С индекса k в последните два израза обозначаваме общия брой на векторите z_{in} (респ. z_{jn}) и s_n, а с β_k регресионните параметри (неизвестни), които ще бъдат обект на статистическо оценяване.

За яснота нека конкретизираме емпирично наблюдаваните изборни решения като една дихотомна променлива:

$$y_{in} = \begin{cases} 1, & \text{ако потребител } n \text{ е избрал алтернатива } i, \\ 0, & \text{ако потребител } n \text{ е избрал алтернатива } j. \end{cases}$$

[2] Тази аксиома е известна и с названието „свойство на еквивалентните разлики“ (вж. Paczkowski, 2016, p. 211).

За удобство бихме могли да изписваме, че $y_{jn} = 1 - y_{in}$.

Нека разгледаме хипотетична изборна ситуация, в която потребителите могат да избират измежду два типа градски транспорт за придвижване между две населени места – влак и личен автомобил. Всяка една от тези две алтернативи има свои специфични атрибутивни характеристики (напр. разходи за едно пътуване, време за едно пътуване, пътническа класа (ако е избран влак). Освен това отделните потребители е възможно да се различават по целта на пътуването, по пол или по социален статус. За осигуряване на проследимост приемаме, че тези конкретни спецификации за продуктовите атрибути (z) и потребителските характеристики (s) са описаните на Таблица 1.

Таблица 1

Хипотетични спецификации на предиктори на дихотомен изборен модел[3]

Параметри на модела	Изборни алтернативи		
	$y_{i=\text{автомобил}}$	$y_{j=\text{влак}}$	
β_1	1	0	Продуктови атрибути
β_2	Разходи за 1 пътуване с автомобил	Разходи за 1 пътуване с влак	
β_3	Времетраене на едно пътуване с автомобил (часове), ако пътуването е със служебна цел, 0 ако е частно.	0	
β_4	Времетраене на едно пътуване с автомобил (часове), ако пътуването не е със служебна цел, 0, ако е частно.	0	
β_5	0	Времетраене на едно пътуване с влак (часове)	
β_6	0	1, ако е избрана първа класа, 0 във всички останали случаи	
β_7	1, ако пътникът е мъж, 0 във всички останали случаи	0	Характеристики на потребителя
β_8	1, ако пътникът е глава на домакинство, 0 във всички останали случаи	0	
β_9	1, ако пътникът има фиксиран час на пристигане, 0 във всички останали случаи	0	

Нека допуснем, че трима респонденти (потенциални пътници) са били поставени в ситуация да избират начин на придвижване. Всяка ситуация е съдържала различни продуктови профили (т.е. различна комбинация от разходи за придвижване и време за пътуване със съответния вид

[3] По аналогия с Биерле (Michel Bierlaire & Lurkin, 2020: 3.3 Maximum likelihood estimation)

транспорт), от които всеки един от тях е бил поканен за избира най-желаната. Данните от подобно допитване е необходимо да се организират в матричен вид (вж. Таблица 2).

Таблица 2

Организация на примерни данни от допитване до трима респонденти

ID	Избор		β_1 =?	β_2 =?	β_3 =?	β_4 =?	β_5 =?	β_6 =?	β_7=?	β_8 =?	β_9=?
1	$y_{i1} = 1$	x_{i1}=	1	40,00	0	1,17	0	0	1	0	0
	$y_{j1} = 0$	x_{j1}=	0	5,00	0	0	2,50	0	0	0	0
2	$y_{j2} = 0$	x_{i2}=	1	28,00	2,00	0	0	0	0	1	1
	$y_{j2} = 1$	x_{j1}=	0	7.80	0	0	1,75	1	0	0	0
3	$y_{i3} = 0$	x_{i3}=	1	45,00	0	2,55	0	0	0	1	0
	$y_{j3} = 1$	x_{j3}=	0	3,20	0	0	2,67	0	0	0	0

Данните за наблюдението върху първия респондент (ID = 1) са поместени на първите два реда от таблицата. Вижда се, че той е предпочел пътуване с автомобил, при разход за едно пътуване от 40 лв., пътувал е частно, като времето за пътуване е 1,17 часа, респондентът е мъж, не е глава на домакинство и няма фиксиран час за пристигане. По същия начин следва да се интерпретират и данните за другите два респондента. Те са избрали да пътуват с влак и продуктовите атрибути и потребителските им характеристики са описани на следващите редове от таблицата.

На практика в този фиктивен сценарий имаме данни от заявения избор на трима респонденти, но са налице 6 наблюдения (по две за всеки респондент). Целта е, на базата на тези данни да бъдат намерени статистически оценки на регресионните коефициентите $\hat{\beta}_1, \dots, \hat{\beta}_K$ (в примера, $K = 9$) с помощта на метода на максималното правдоподобие.

Методът на максималното правдоподобие има някои интересни свойства. Например с увеличаване на размера на извадката (в случая 6 наблюдения) оценките на параметрите на модела постепенно конвергират до техните „истински“ стойности. Освен това оценките на параметрите са асимптотично нормално разпределени (т.е. за валидирането им са приложими всички стандартни параметрични статистически тестове). Самата процедура цели определяне стойността на неизвестните параметри на модела по такъв начин, че съвместната вероятност на наблюдаваните решения, предсказвани от модела, да бъде възможно най-голяма. По терминологията на метода „съвместната“ вероятност е прието да се нарича правдоподобност. Правдоподобността е функция от параметрите на модела.
С примерните данни от Таблица 2 правдоподобността може да се получи по следния начин:

- Първият респондент е избрал „автомобил“ като алтернатива за придвижване. Неговият избор следва да се предскаже от модела с вероятност $P_1(i, \{i, j\})$.

- Вторият респондент е избрал алтернативата „влак“. Предсказаната от модела вероятност отбелязваме с $P_2(j, \{i, j\})$.
- Третият респондент също е избрал алтернативата „влак“ и предсказаната от модела вероятност следва да се обозначи с $P_3(j, \{i, j\})$.

Съвместната вероятност (т.нар. „правдоподобност” $\mathcal{L}^*$), с която моделът би следвало да предскаже едновременно трите емпирично наблюдавани изборни решения, може да се изрази като произведение:

$$\mathcal{L}(\beta_1, \dots, \beta_9) = P_1(i) \cdot P_2(j) \cdot P_3(j).$$

Ако значенията на регресионните параметри β_k са нулеви ($k = 1, \dots, K$), правдоподобността ще бъде равна на:

$$\mathcal{L} = \frac{e^0}{e^0+e^0} \cdot \frac{e^0}{e^0+e^0} \cdot \frac{e^0}{e^0+e^0} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = 0{,}125.$$

Ако регресионните параметри β_k получат ненулеви стойности (напр. 3,04;-0,05;-2.66;-2.22;-0.57;0.96;-0.85;0.38;-0.62), то правдоподобността започва да нараства:

$$\mathcal{L} = 0{,}325 \cdot 0{,}970 \cdot 0{,}943 = 0{,}297.$$

Целта обаче е да се намери тази комбинация от оценки на параметрите β_k, при която $\mathcal{L}$ достига максимално възможна стойност. Ако обобщим ситуацията, приемайки извадка с достатъчно голям брой случайно наблюдавани изборни ситуации на N респонденти, то отново правдоподобността ще представлява произведение от вероятностите $P_n(.)$ на отделните наблюдения. Така функцията на правдоподобност изглежда може да се изрази в общ вид (Bierlaire, 2015):

$$\mathcal{L}(\beta_1, \beta_2, \dots, \beta_K) = \prod_{n=1}^{N} P_n(i)^{y_{in}} P_n(j)^{y_{jn}},$$

в който $P_n(i)$ и $P_n(j)$ са функции на $\beta_1, \dots, \beta_K$, а всеки фактор представлява изборната вероятност на избраната алтернатива, т.е.:

$$P_n(i)^{y_{in}} P_n(j)^{y_{jn}} = \begin{cases} P_n(i), \text{ако } y_{in} = 1 \text{ и } y_{jn} = 0 \\ P_n(j), \text{ако } y_{in} = 0 \text{ и } y_{jn} = 1 \end{cases}$$

Функцията на правдоподобност е сравнително трудна за анализиране поради сумата от произведенията на отделните вероятности. Много по-лесен за манипулация е нейният логаритмуван вариант, тъй като логаритъм от произведение на множители е равен на сумата от логаритмуваните му елементи. Логаритмувайки двете страни на функцията $\mathcal{L}$, получаваме:

$$LL(\beta_1, \dots, \beta_K) = \sum_{n=1}^{N} \big(y_{in} \ln P_n(i) + y_{jn} \ln P_n(j)\big).$$

$\mathcal{L}$ е възможно да заема стойности в диапазона [0,1] и при големи извадки тези стойности са относително ниски. Логаритмуваният й вариант LL обаче има допустим интервал от стойности в диапазона $(-\infty, 0)$. Търсенето на оптималните оценки на параметрите β_k е по същество оптимизационен проблем, състоящ се в изследване на функцията LL за максимум:

$$\max LL(\hat{\beta}_k) = LL(\hat{\beta}_1, \dots, \hat{\beta}_K),$$

където $\hat{\beta}$ представлява вектор с елементи, съставени от оценките $\hat{\beta}_1, \dots, \hat{\beta}_K$. Решаването на оптимизационния проблем предполага проверка на условията за оптималност, а именно – частните първи производни на функцията спрямо неизвестните β_k да са равни на 0:

$$\frac{\partial LL}{\partial \beta_k}(\hat{\beta}) = \sum_{n=1}^{N}\left(y_{in}\frac{\partial P_n(i)/\partial \beta_k}{P_n(i)} + y_{jn}\frac{\partial P_n(j)/\partial \beta_k}{P_n(j)}\right) = 0, \text{като } k = 1, \dots, K.$$

Изразът $\partial LL(\hat{\beta})/\partial \beta$ представлява вектор на първите частни производни на логаритмуваната функция на правдоподобност спрямо неизвестните параметри. За да бъдат намерени оценките на $\hat{\beta}$, удовлетворяващи по-горното условие за оптималност, е необходимо да се приложи итеративна процедура. Тя започва със случайно зададени стартови или с нулеви стойности на неизвестните параметри. Прави се проверка, дали всички първи частни производни на функцията са равни на нула. Ако това условие не е изпълнено, стартовите стойности на неизвестните се променят с малка стъпка и отново се прави за приближението на първите производни към 0. Процесът се повтаря итеративно до момента, в който не е възможно „подобрение“ на достигнатия максимум. Стъпката за изменение на стойностите на неизвестните параметри се контролира от анализиращия.

Описаният процес се реализира най-често чрез числов алгоритъм, известен като метод на Нютон (или Нютон-Рафсън). Този алгоритъм използва поредица от последователни все по-точни итеративни приближения ℓ до достигане на търсената точност на решението. На всяка итерация ℓ се съставя квадратичен модел на логаритмуваната функция на правдоподобие върху текущата итерация $\beta^{(\ell)}$. Моделът се съставя така, че стойността му и неговите първи и втори частни производни да са еднакви при $\beta^{(\ell)}$:

$$m(\beta; \beta^{(\ell)}) = LL(\beta^{(\ell)}) + (\beta - \beta^{(\ell)})^T \nabla LL(\beta^{(\ell)}) + \\ + \frac{1}{2}(\beta - \beta^{(\ell)})^T \nabla^2 LL(\beta^{(\ell)})(\beta - \beta^{(\ell)}),$$

където $\nabla LL(\beta^{(\ell)})$ е граидент, представляващ вектор с първите частни производни на логаритмуваната функцията на правдоподобност, изведени при итерация $\beta^{(\ell)}$, а $\nabla^2 LL(\beta^{(\ell)})$ е матрицата с вторите частни производни, изведени на същата итерация. Посоката на всяка последваща итерация се извършва, когато стойностите на параметрите максимализират квадратичния модел:

$$\beta^{(\ell+1)} = \beta^{(k)} - \nabla^2 LL(\beta^{(\ell)})^{-1} \nabla(\beta^{(\ell)}).$$

В числов вид това се получава чрез решаването на системата от линейни уравнения $\nabla^2 LL(\beta^{(\ell)})d = -\nabla(\beta^{(\ell)}$, за да се установи посоката на изменение d, след което се изчислява $\beta^{(\ell+1)} = \beta^{(\ell)} + d$. Процедурата продължава, докато градиентът се доближи достатъчно плътно до нулата (което зависи от степента на прецизност, задавана от анализатора или зададена по подразбиране в съответния софтуерен код). Типичен минимален праг на прецизност е 10^{-6} (за повече подробности вж. напр. Bierlaire, M. 2015).

След оценяването на параметрите на дихотомния изборен модел възниква следващият важен изследователски въпрос – доколко оцененият модел е статистически валиден, за да може да се използва за прогнозни цели. Проверката на **валидността** на оценения модел е възможно да бъде извършена чрез обследване на стандартните грешки σ_k на оценките на параметрите β_k, чрез t-тест за статистическа значимост на оценените параметри и респ. на базата на статистическата значимост на оценките при ниво на доверие напр. 95%. Стандартните грешки на оценките на β_k могат да се изчислят като:

$$\sigma_k = \sqrt{\hat{\Sigma}_\beta(k,k)},$$

където $\hat{\Sigma}_\beta(k,k)$ представлява k-тия елемент на диагонала на матрицата $\hat{\Sigma}_\beta$.

При наличието на изчислени стандартни грешки може да се намери емпиричната t-стойност за всяка оценка $\hat{\beta}_k$:

$$t_k = \frac{\hat{\beta}_k}{\sigma_k}.$$

Този критерий се използва обикновено за проверка на нулевата хипотеза, гласяща, че истинската стойност на параметъра β_k е нула. Тази хипотеза може да бъде отхвърлена с 95% сигурност, ако $|t_k| \geq 1{,}96$.

При наличие на емпиричните стойности на t_k е целесъобразно и удобно да се изчисли нивото на статистическа значимост p:

$$p_k = 2(1 - \Phi(t_k)),$$

където $\Phi(\cdot)$ представлява кумулативната вероятностна функция на едномерното стандартно нормално статистическо разпределение.

Нивото на значимост p дава същата информация, както и t-критерия, но представена по различен начин. То показва вероятността, емпиричната t-стойност да е поне толкова голяма (в абсолютно изражение), колкото тази, получена при нулевата хипотеза, т.е. че $\beta_k = 0$. Интерпретацията е, че нулевата хипотеза може да бъде отхвърлена с ниво на валидност $1 - p_k$. Обикновено за критично се приема нивото от 0,05.

Освен тези стандартни статистики за проверка на статистическата значимост на оценките на коефициенти $\hat{\beta}_k$, необходимо е да се оцени и адекватността на оценения модел като цяло. За разлика от линейната регресия, тук могат да се използват различни критерии за адекватност. Нито един от тях обаче не може да се използва самостойно. Критериите за адекватност на изборните модели могат да бъдат използвани само при сравняването на два или повече оценени модела.

Най-общият критерий за адекватност на оценен изборен модел е самата логаритмувана правдоподобност. Обикновено тази оценка се сравнява с оценката на някакъв друг отправен сравним модел (напр. с т.нар. „нулев“ модел, в който всички оценки на параметрите се приемат за нули). Ако приемем например един съвсем тривиален нулев модел, в който вероятността за избор на всяка от две възможни алтернативи е 0,5 (т.е. 50%)

$$P_n(i\{i,j\}) = P_n(j\{i,j\}) = \frac{1}{2}$$

и предположим, че разполагаме с N наблюдения, то логаритмуваната правдоподобност на този модел би била:

$$LL(0) = \log\left(\frac{1}{2^N}\right) = -N\log(2).$$

Тази оценка би могла да бъде използвана за извеждане на статистически коефициент на правдоподобност λ_{LR}, съдържащ разликата между нулевия модел $LL(0)$ и оценения модел $LL(\hat{\beta})$ (Ben-Akiva & Lerman, 1985, p. 166):

$$\lambda_{LR} = -2\big(LL(0) - LL(\hat{\beta})\big).$$

Коефициентът на правдоподобност се използва за проверка на нулевата хипотеза H_0, гласяща, че оцененият модел е еквивалентен на нулевия модел (т.е. на модела с еднаквите изборни вероятности). Приема се, че $-2((LL0) - LL\big(\hat{\beta}\big))$ следва приближено известното χ^2 разпределение със степен на свобода K.

Дотук разглеждахме изборен модел, в който са включени само две алтернативи за избиране. Изхождайки от хипотезата, че грешките на модела следват статистическо разпределение на екстремните стойности на Гумбел, обосновахме и изведохме дихотомен логит модел за неговото оценяване. Аналитично описахме как неизвестните параметри на модела биха могли да бъдат намерени по статистически път с данни от извадка от наблюдения на изборни ситуации. Предложихме като подходящ метод за оценяване на параметрите на моделът методът на максималното правдоподобие, реализуем с числовия алгоритъм на Нютон-Рафсън. В следващата точка от изложението ще се опитаме да разширим обекта на изследване, разглеждайки изборни ситуации, включващи повече от две алтернативи за избор.

2.2. Моделиране на избора при множество алтернативи

В реалните пазари потребителите попадат най-често в ситуации, които предполагат вземане на решение за избор измежду повече от две алтернативи за покупка. В този раздел се представят процесите на идентифициране и оценяване на логит модели, включващ повече от две алтернативи. Подобни модели ще наричаме мултиномиални модели или просто модели на многочленен избор. Особено внимание се обръща върху привеждането на поведенческите хипотези в математически модели.

2.2.1. Общ вид на мултиномиалния модел

При извеждането на модел на избора, допускащ наличието на повече от две изборни алтернативи, приехме хипотезата, че всеки потребител n е изправен пред изборно множество C_n, състоящо се от J_n продуктови алтернативи, т.е. $C_n = \{1, \dots, J_n\}$. Както вече бе обосновано в предходните раздели, функцията на полезност U_{in} на всяка продуктова алтернатива i за всеки отделен потребител n се състои от детерминистичен и стохастичен компонент $U_{in} = V_{in} + \varepsilon_{in}$. Допускаме, че стохастичните компоненти ε_{in} са (1) независими една от друга както между продуктовите алтернативи, така и между отделните потребители и (2) следват статистическото разпределение на екстремните стойности на Гумбел, което е с едни и същи параметри за всеки потребител и за всяка продуктова алтернатива, а именно – нулева очаквана стойност и един и същи мащаб (дисперсия μ):

$$\varepsilon_{in} \sim EV(0, \mu)$$

Тези допускания се обобщават в аксиомата за независимост от ирелевантните алтернативи. Изхождайки от тях, вероятността за избора P на i-тата алтернатива от страна на i-тия потребител от достъпното му изборното множество C_n може да се изрази като[4]:

$$P(i|C_n) = \Pr\left(V_{in} + \varepsilon_{in} \geq \max_{j=1,\dots,J_n} V_{in} + \varepsilon_{jn}\right)$$

Идеята е, този модел да се разглежда като дихотомен логит, тъй като вече изведохме неговата спецификация в предходния раздел. Моделът показва, че за да бъде избрана продуктовата алтернатива i от даден потребител, то тя трябва да има полезност, по-голяма от полезностите на всички останали алтернативи в изборното множество C_n. Ако отбележим с $C_n \backslash \{i\}$ множеството на останалите алтернативи (без i), изследователят не би могъл да знае предварително, коя/кои от тях има най-висока полезност, но би могъл да допусне, че нейната полезност U_n^* за конкретния потребител n би трябвало да бъде най-висока в сравнение с тези на останалите продукти, съставящи изборното множество $C_n \backslash \{i\}$:

$$U_n^* = \max_{j \in C_n \backslash \{i\}} U_{in} = \max_{j \in C_n \backslash \{i\}} (V_{in} + \varepsilon_{in})$$

Оттук и съобразно изведения вече модел на дихотомния логит мултиномиалният модел на избора би могъл да бъде представен като съдържащ само две изборни алтернативи:

$$P(i|C_n) = \Pr(U_{in} \geq U_n^*)$$

Следователно, за да бъде изведен операционален модел за предсказване на потребителския избор, е необходимо да знаем разпределението на U_n^*.

Изхождайки от разпределението на екстремните стойности на Гумбел и приемайки, че всички грешки са в съответствие с аксиомата за

[4] По аналогия с Bierlaire & Lurkin, (2020): 5.1 Derivation of the logit model.

независимост от ирелевантните алтернативи, разпределението на U_n^* има следните параметри[5]:

$$U_n^* \sim EV\left(\frac{1}{\mu} \ln \sum_{j \in C_n \setminus \{i\}} e^{\eta V_{jn}}, \mu\right).$$

Полезността обаче съдържа както детерминистична, така и стохастична компонента $U_n^* = V_n^* + \varepsilon_n^*$. Отчитайки, че очакваната стойност на $V_n^* = \frac{1}{\mu} \ln \sum_{j=2}^{J_n} e^{\mu V_{jn}}$ и приемайки $\varepsilon_n^* \sim EV(0, \mu)$, то дихотомният логит модел би добил вида:

$$P(i|C_n) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_n^*}}.$$

Тъй като $e^{\mu V_n^*} = e^{\ln \sum_{j \in C_n \setminus \{i\}} e^{\mu V_{jn}}} = \sum_{j \in C_n \setminus \{i\}} e^{\mu V_{jn}}$, то окончателният вид на мултиномиалния логит модел е следният (Bierlaire & Lurkin, 2020: 5.1 Derivation of the logit model):

$$P(i|C_n) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + \sum_{j \in C_n \setminus \{i\}} e^{\mu V_{jn}}} = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}}.$$

Видно е, че мултиномиалният логит е разширение на дихотомния логит модел, като сумата в знаменателя сега включва всички алтернативи от изборното множество.

Специфично при емпирично оценяване на този модел е обстоятелството, че параметърът на мащаба (дисперсията) μ е невъзможно да бъде идентифициран и оценен на базата на емпирични данни от наблюдения. Например, ако $V_{in} = \sum_k \beta_k x_{ink}$, то влиянието на мащаба в модела би било $\mu V_{in} = \sum_k \mu \beta_k x_{ink}$. Следователно само произведението $\mu \beta_k$ може да бъде идентифицирано и оценявано чрез данни. По тази причина е общоприето, параметърът μ да се приема за равен на 1 и да се оценява само коефициентът β_k. Фактът, че параметърът на мащаба не може да бъде идентифициран, не означава, че не съществува. Например, ако стойността на μ клони към нула, т.е. когато дисперсията на грешките клони към безкрайност, детерминистичната част от полезността V_{in} вече не играе роля и моделът би предсказал една и съща вероятност за всяка алтернатива от изборното множество. И обратното, ако стойността на μ клони към безкрайност, т.е. дисперсията на грешките се доближава до нула, моделът се превръща в изцяло детерминистичен.

2.2.2. Оценяване на мултиномиалния модел

Вече подчертахме нееднократно, че общият вид на модела на случайната полезност се състои от два компонента, детерминистичен и сто-

[5] Едно от свойствата на разпределението на екстремалните стойности гласи, че ако $\varepsilon_i \sim EV(\eta_i, \mu)$ при $i = 1, \dots, J$ и ако ε_i са независими помежду си и с един и същ мащаб (дисперсия) μ, то $\varepsilon = \max_{i=1,\dots,J} \varepsilon_i \sim EV(\eta, \mu)$, където очакваната стойност η на това разпределение е равна на $\frac{1}{\mu} \ln \sum_{i=1}^{J} e^{\mu \eta_i}$.

хастичен, и за всяка изборна алтернатива i, принадлежаща на изборното множество C_n на потребител n, е в сила:

$$U_{in} = V_{in} + \varepsilon_{in}.$$

Оттук при емпиричното оценяване на мултиномиални логит модели стоят три основни въпроса – как да бъдат специфицирани (1) изборните множества C_n, (2) стохастичните компоненти ε_{in} и (3) детерминистичните компоненти V_{in}.

Спецификация на изборните множества

С понятието „изборно множество“ обозначихме списъка от алтернативи, измежду които потребителят е възможно да избира. В този смисъл различаваме универсално и индивидуално изборно множество. Под универсално изборно множество разбираме всички потенциални продуктови алтернативи, които хипотетично биха могли да бъдат избирани от целевия пазар. Универсалното изборно множество би могло да бъде ограничено до множество на релевантните продуктови алтернативи, които представляват изследователски интерес за анализиращия (напр. всички конкурентни марки в конкретна продуктова категория или всички възможности за придвижване в градска среда). От гледна точка на отделния потребител не всички алтернативи биха могли да бъдат обект на избор и решение за покупка. Например, някои от алтернативите в универсалното изборно множество може да не са достъпни и/или неизвестни за конкретен потребител (напр. някои домакинства не разполагат с автомобил, свидетелство за правоуправление, велосипед и/или удобен достъп до метро). Поради тези причини с понятието „индивидуално изборно множество“ отбелязваме само набора от алтернативи, измежду които избира конкретният потребител. Това означава, че всеки отделен потребител би могъл да разглежда отделен, ограничен набор от алтернативи за решение за покупка. Индивидуалното изборно множество на отделния потребител го отбелязахме вече с C_n.

Идентифицирането на индивидуалните изборни множества за всеки потребител е сериозно предизвикателство за емпиричното наблюдение, тъй като е необходимо да се моделират информираността, възприятията и нагласите на всеки отделен респондент. Обикновено се използва интервю с адаптивен избор (наричано „адаптивен изборен конджойнт анализ“), което е интерактивно преживяване. Въпросникът в това интервю е съобразен с предпочитанията, възприятията и мнението на всеки респондент. Стремежът е, чрез индивидуализирано интерактивно анкетиране да се проучи в дълбочина структурата на вземане на решения от всеки респондент (Orme & Johnson, 2007) и да се идентифицира индивидуалното му изборно множество C_n чрез систематични правила. Например в индивидуалното изборно множество на респондент n ще попадне алтернативата „автомобил“, ако съответният респондент е заявил, че разполага с валидно свидетелство за управление и домакинството притежава автомобил. Други правила биха могли да бъдат наличието на предлагане на конкретна марка в обсега на клиента или субективен праг на дистанцията, до която той би могъл да се придвижва пеш до желана ежедневна дестинация. Информацията за тези „правила“ е необходимо да бъде регистрирана чрез индикатори в масива с изходни данни за всеки отделен респондент, попадащ в наблюдението. Например, за всеки потребител (респондент) n и всяка алтернатива i в изборното множество може да се създаде дихотомен индикатор δ_{in}, който получава значение 1, ако съответната алтернатива попада в изборното множество на респондента C_n и 0, ако не попада, т.е.:

$$\delta_{in} = \begin{cases} 1, \text{ако } i \in C_n, \\ 0, \text{ако } i \notin C_n. \end{cases}$$

Използвайки тези индикатори като променлива в общия вид на изборния модел, то той ще придобие следния вид:

$$P_n(i|C_n) = P_n(i|\delta_n, C) = \Pr(U_{in} + \ln\delta_{in} \geq U_{jn} + \ln\delta_{jn}),$$

в който с C обозначаваме универсалното изборно множество.

В случай, когато δ_{in} е 1, то $\ln\delta_{in} = 0$, което ни връща към оригиналната оценка за полезност U_{in}. В случай че δ_{in} е 0, $\ln\delta_{in} \rightarrow -\infty$ и следователно прилежащата полезност не би могла да бъде възможно най-високата. По този начин при кодирането на модела, в зависимост от изследователския контекст, е възможно, той да бъде представен или като функция от универсалното изборно множество, или като функция от индивидуалните изборни множества.

Спецификация на детерминистичния компонент

Детерминистичните компоненти V_{in} на функцията на полезността U_{in} представляват най-съществената част от модела на полезността U_{in}, тъй като чрез него се дава математическо изражение на теоретичните хипотези за потребителския избор. По същество V_{in} представлява функция, в която независимите променливи се подразделят на две категории – първата, това са продуктовите атрибути на алтернативите, формиращи индивидуалните изборни множества C_n, а втората – характеристиките на самите потребители, които изследователят допуска, че обясняват потребителския избор. Ако със z_{in} отбележим вектора с атрибутивните характеристики на изборните алтернативи i за потребител n, а със s_n изразим вектора със социално-демографски и/или психографски характеристики на същия този потребител n, то формално можем да запишем, че:

$$V_{in} = V(z_{in}, s_n).$$

Първият набор от променливи z_{in} има два индекса, като те варират както при потребителите, така и при самите изборни алтернативи. Вторият набор от променливи S_n варира само при потребителите и затова имат само един индекс i.

В общия случай се приема, че формата на тази функция е линейна по отношение на съставящите я параметрите, т.е. V_{in} е линейна комбинация от наборите променливи x_{in} и S_n. Ако за удобство приемем, че $x_{in} = h(z_{in}, s_n)$, то детерминистичният компонент може да се изрази като:

$$V_{in} = V(z_{in}, s_n) = V(x_{in}) = \sum_{k=1}^{K} \beta_k (x_{in})_k.$$

В случая, с индекса k обозначаваме общия брой на векторите z_{in} и s_n, а с β_k – регресионните параметри (неизвестни), които подлежат на статистическо оценяване с помощта на емпирични данни.

Презумпцията за „линейност“ на пръв поглед изглежда наивно и нереалистично отразяване на реалността. Винаги може да се допусне, че функционалните закономерности, които „управляват“ сложните феномени в поведение на потребителите, са по-скоро нелинейни. Математическото „опростяване“ обаче, освен че улеснява числовото оценяване на модела, не е чак толкова рестриктивно.

Спецификация на стохастичния компонент

За стохастичните компоненти ε_{in} в мултиномиалния логит модел се изхожда от същата хипотеза, както и при дихотомния логит. Допуска се, че ε_{in} са случайни променливи, независими както от отделните алтернативи i, така и от отделните потребители n. Стохастичните компоненти следват идентични гумбелови разпределения на екстремните стойности с параметър на математическото очакване $\eta = 0$ и дисперсия (мащаб) μ. Идентични означава, че параметрите η и μ са едни и същи за всички потребители n, както и за всички изборни алтернативи i. При извеждането на общия вид на мултиномиалния логит модел вече отбелязахме, че параметърът μ е невъзможно да бъде идентифициран от емпирични данни и по тази причина, при оценяването на модела неговата стойност се приема за 1.

Спецификация на хетерогенността

В реалния свят хората се държат по различен начин. Всеки един потребител би могъл да има нещо специфично в поведението си при избиране на конкретен продукт или услуга. Пазарите са съставени от хетерогенни множества от купувачи и това би следвало да бъде обхващано от моделите на избора. Темата за хетерогенността между потребителите е фундамент в маркетинговата концепция, тъй като осигурява основа за сегментиране на пазара, таргетиране, продуктово позициониране и масово персонализиране. Съществуват множество изследвания, насочени към включването на хетерогенността в моделите за избор на марка. Повечето проучвания в тази област обаче са фокусирани върху разликите в предпочитанията или вкусовете на потребителите. За разлика от това ограничено внимание се отделя на възможността, потребителите да се различават и в процеса, който следват, докато правят избор. Неотчитането на потребителска хетерогенност може да доведе до неправилно тълкуване на пазарната структура и пазарните сегменти. В тази връзка един от сериозните приноси върху начините за интегриране на потребителска хетерогенност при моделиране на избора правят Камакура (Kamakura et al., 1996), както и Десарбо (Desarbo et al., 1997). Те разработват варианти на модели за избор, които едновременно идентифицират потребителските сегменти въз основа на предпочитания, реакцията към маркетинговия микс и процесите на избор. Техническите детайли оставяме извън ограничения обхват на настоящата студия.

3. Валидност на моделите на дискретния избор

Многообразието от варианти и спецификации на модели на избора с помощта на мултиномиален логит предполага сериозно предизвикателство за статистическото им оценяване. За да бъдат използвани за аналитични цели и за основание за вземане на конкретни маркетингови решения, оценените модели трябва да бъдат подлагани на прецизна проверка за валидност. Проверката за валидност може да бъде както неформална (логическа), така и статистическа. В тази точка ще бъдат разгледани накратко подходите за проверка на статистическата и предиктивната валидност на изборните модели.

3.1. Проверка за адекватност на модела

Към момента не е известен универсален критерий, с помощта на който да бъде оценявана общата статистическа адекватност на логит моделите на избора. Единственият начин за проверка на „качеството“ на оценен модел е чрез сравняването му с някакъв друг референтен модел чрез теста за правдоподобност.

Тестът за правдоподобност оценява степента на съответствие на два конкурентни статистически модела въз основа съотношението на тяхната правдоподобност. В общия случай оценяваният модел се сравнява с някакъв референтен (ограничен) модел. За нулева хипотеза се приема допускането, че правдоподобността на ограничения модел не се различава статистически значимо от тази на оценявания (неограничен) модел. По този начин тестът за правдоподобност прави проверка, дали съотношението на правдоподобностите се различава от единица (или алтернативно, дали натуралният логаритъм на съотношението на правдоподобностите се различава статистически значимо от нула (King, 2010, p. 84)). Чрез този тест е възможно да се изследват различни версии на спецификация на даден модел, като се въведат линейни ограничения върху параметрите.

Обикновено за референтен се приема модел, в който са наложени някакви ограничения. Най-лесният начин е да се приеме модел, в който вероятностите за избор на всяка разглеждана алтернатива i при всеки потребител n са равни, т.е. $\beta_{ink} = 0, \forall\, i, n, k$. Това по същество означава, че потребителите избират напълно случайно измежду всички възможни алтернативи. Функцията на правдоподобност на оценявания (неограничен) модел би могла да се изрази в общ вид по познатия вече начин като $\mathcal{L}(\hat{\beta})$.

Тъй като ограниченият модел предполага еднаква изборна вероятност, т.е. $P_{in} = \frac{1}{J_n}, \forall i \in C_n, \forall\, n$, то неговата функция на правдоподобност има вида $\mathcal{L}(0) = -\sum_{n=1}^{N} \log(J_n)$. Оттук като критерий за приемане или отхвърляне на нулевата хипотеза може да се използва т.нар. коефициент на правдоподобност λ_{LR}, който има следния вид (Ben-Akiva & Lerman, 1985, p. 166):

$$\lambda_{LR} = -2(\mathcal{L}(0) - \mathcal{L}(\hat{\beta}) \sim \chi_K^2.$$

Приема се, че λ_{LR} следва χ^2 статистическо разпределение със степен на свобода K. В случая степента на свобода представлява разликата между броя на параметрите на двата сравнявани модела. Тъй като ограниченият модел, ползван като референтен, няма параметри, тъй като ги приехме за нулеви, то K е равно на броя на оценяваните параметри в неограничения модел.

Сравняването с нулев референтен модел с равни вероятности не е особено полезно, тъй като подобен модел не може да предсказва надеждно изборната вероятност. Затова обикновено се търси друг вариант за „ограничен“ модел, който да се ползва като референтен. Възможно е да се състави модел, който да съдържа само специфични за продуктовите алтернативи константи. По този начин се ограничават всички коефициенти, с изключение на споменатите константи. При тези ограничения изборната вероятност ще възпроизвежда на практика пазарния дял на база данните от извадката, тъй като вероятността за избор на продуктова алтернатива i представлява дела на респондентите от извадката, които са посочили (избрали) тази алтернатива, т.е. $P_{in} = N_i/N\ \forall\, i \in C, \forall\, n$. Функцията на правдоподобност на този ограничен модел може да се изрази като $\mathcal{L}(c) = \sum_{i=1}^{J} N_i \log(\frac{N_i}{N})$, като припомняме, че J изразява броя на продуктовите алтернативи. В този случай критерият за приемане или отхвърляне на нулевата хипотеза (т.нар. коефициент на правдоподобност λ_{LR}) ще има вида:

$$\lambda_{LR} = -2\left(\mathcal{L}(c) - \mathcal{L}(\hat{\beta})\right) \sim \chi_d^2.$$

С d последния израз е отбелязана степента на свобода, която е равна на разликата между броя K на параметрите на неограничения модел и броя на параметрите на ограничения модел $J - 1$, т.е. $d = K - J + 1$.

Възможно е да бъде дефиниран и трети вариант на рестриктивния модел, като например той да съдържа коефициенти само за т.нар. общи променливи, които влияят върху полезността и респ. върху вероятността за избор на продуктовите алтернативи по един и същ начин. При този сценарий коефициентът на правдоподобност ще бъде:

$$\lambda_{LR} = -2\left(\mathcal{L}(\hat{\beta}_G) - \mathcal{L}(\hat{\beta})\right) \sim \chi_d^2,$$

като $d = K - K_G$. $\mathcal{L}(\hat{\beta}_G)$ изразява функцията на правдоподобност на ограничения само до общите параметри модел, а $\mathcal{L}(\widehat{(\beta})$ е функцията на правдоподобност на неограничения модел.

Възможни са и други варианти за прилагане на теста за правдоподобност, отчитащи например хетерогенността на потребителите, или отчитайки нелинейни зависимости между променливите на модела. Независимо от използвания вариант, получената емпирична стойност на коефициента на правдоподобност λ_{LR} се сравнява с теоретичната χ^2 стойност, отчетена при съответната степен на свобода d и приемлив праг за α-грешка (напр. 0,05). Ако емпиричната стойност е по-висока от теоретичната, или респ. ако $p \leq 0{,}05$, нулевата хипотеза, гласяща, че правдоподобността на сравняваните модели е еднаква, следва са се отхвърли и да се направи извод, че неограниченият модел е статистически валиден.

Бен-Акива и Лерман разработват и предлагат на базата на теста за правдоподобност т.нар. индекс на правдоподобност ρ^2 (Ben-Akiva & Lerman, 1985, p. 167):

$$\rho^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)}.$$

Интерпретацията на ρ^2 наподобява тази на коефициента на девиация R^2 при регресионния анализ. Ако оценяваният модел има същата логаритмична правдоподобност като нулевия модел, то $\rho^2 = 0$. Ако оцененият модел възпроизвежда перфектно емпиричните данни, т.е. ако $LL(\hat{\beta}) = 0$, то $\rho^2 = 1$.

Съществена слабост на този индекс е, че той по подобие на R^2 монотонно нараства с увеличаване броя на променливите K, дори и ако тези променливи не обясняват нищо. По тази причина се изчислява и негова „коригирана“ версия, имаща вида:

$$\bar{\rho}^2 = 1 - \frac{LL(\hat{\beta}) - K}{LL(0)}.$$

Както бе споменато по-горе, самата стойност не може да бъде интерпретирана и трябва да се използва само за сравняване на два модела. По-специално, за разлика от линейната регресия, възможно е да има добър модел с ниска ρ^2 стойност и лош модел с висока ρ^2 стойност.

Тестът за правдоподобност би могъл да се прилага при различни спецификации на изборния модел, но той не е универсален. Възможни са обаче ситуации, при които той е неприложим. Например, ако референтният модел не представлява линейна рестрикция на неограничения модел, тестът за правдоподобност не е допустим. В този случай е възможно да бъде приложен тестът на Кокс (Cox, D. R. 1962; Cox, David R. 1961). При теста на Кокс основната идея е, от два изходни модела (напр. модел 1 и модел 2), които не подлежат на директно сравняване с теста за правдоподобност, да се състави трети, композитен модел С. Двата изходни модела се разглеждат като линейно ограничени модели на композитния. След това за всяка наредена двойка „ограничен модел – композитен модел“ се прилага тестът за правдоподобност. При тези сравнения са възможни четири алтернативни резултата, в зависимост от това, дали след проверката на статистическите хипотези се приема или отхвърля H_0 (Таблица 3).

Таблица 3

Възможни заключения при теста на Кокс

Модел „С“ срещу модел 1	Модел „С“ срещу модел 2	Заключение
Модел 1 не се отхвърля	Модел 2 се отхвърля	Предпочитан е модел 1
Модел 1 се отхвърля	Модел 2 не се отхвърля	Предпочитан е модел 2
Модел 1 се отхвърля	Модел 2 се отхвърля	Необходимо е по-добър модел
Модел 1 не се отхвърля	Модел 2 не се отхвърля	Необходим е друг тест

Източник: (Fisher & McAleer, 1979, p. 148).

Ограничение при теста на Кокс е обстоятелството, че при съставянето на композитен модел, съчетаващ параметрите и променливите на двата изходни модела, е възможно да се получи сложен модел с огромен брой параметри, чието статистическо оценяване би могло да е проблематично. Възможно решение на този проблем предлагат Дейвидсън и МакКинън, разработвайки т.нар. J-тест (Davidson & MacKinnon, 1981).

3.2. Проверка на предиктивната способност на модела

Едно от основните приложения на моделите за избор е за предсказване бъдещото поведението на потребителите. Затова е важно, преди използването на модела като база за вземане на маркетингови решения да се провери дали ще е в състояние да предскаже това, което вече сме наблюдавали. Прилагайки метода на максималното правдоподобие, би следвало вече да разполагаме с оценки на параметрите на модела, които репродуцират възможно най-адекватно наблюдаваните изборни решения. Абсолютно точното възпроизвеждане на избора на всеки отделен респондент от извадката едва ли е възможно и реалистично, поради стохастичния характер на наблюдаваните процеси на индивидуално равнище. Затова е препоръчително да се обследват с особено внимание тези случаи (понякога наричани „отдалечени“ или „нетипични“), при които моделът демонстрира ниска предиктивна точност, и да се разбере евентуално причината за ниското качество на предсказване. Процедурата за обследване на нетипичните случаи предполага (1) оценяване на модела с данните от извадката, (2) изчисляване на предсказаните изборни вероятности с помощта на оценения модел, (3) сортиране и идентифициране на тези случаи, които са грешно предсказани вероятности, респ. при които предсказаната изборна вероятност е най-ниска и (4) тестване на чувствителността на модела от нетипичните случаи (напр. чрез премахването им от масива с данни от извадката). Ако премахването им оказва съществен ефект върху логаритмуваната правдоподобност на модела $LL(\hat{\beta}_k)$, то това е симптом, който заслужава внимание. Възможни са различни причини, провокиращи ниска предиктивна вероятност при ограничен брой случаи. Някои от тях са тривиални, напр. грешки в кодирането. Този тип грешки могат да бъдат отстранени сравнително лесно. Други обаче могат да се „крият“ в грешна спецификация на модела (напр. пропускане на някакви важни независими социодемографски променливи, пренебрегване на хетерогенността между потребителите и/или допускане на нелинейност при съставянето на модела). Този тип причини са сравнително трудни за недвусмислена идентификация и се отстраняват по-скоро на базата на опита или интуицията на изследователя, отколкото въз основа на някакви рационални правила.

В определени случаи е възможно да се наблюдава и някакво девиантно, необяснимо отклонение в поведението на потребителите, което не почива на рационална база. При подобни сценарии не бива прибързано да се отстраняват данните от тези наблюдения, тъй като това би довело до свръхизглаждане на модела. В края на краищата моделът трябва да показва как в действителност се държат хората, а не как трябва да се държат. Не бива да се забравя, че ролята на стохастичния компонент ε_{in} в модела е именно да отразява това несъответствие между очакваната от изследователя поведенческа хипотеза и действителното поведение.

След внимателното обследване на нетипичните случаи е целесъобразно да се пристъпи към проверка на способността на модела да предсказва нови, „непознати“ за него случаи. Този тип тестване заимства идеи от областта на машинното обучение и може да се осъществи по два начина – чрез т.нар. „валидиране извън извадката“ или чрез т.нар. „кръстосано валидиране“. При валидиране извън извадката е необходимо, изходните данни да се подразделят на случаен принцип на две подизвадки в съотношение 80:20, 70:30 или друга произволна пропорция. По-голямата подизвадка се нарича „оценъчна“ или „обучаваща“ и с нея се оценяват параметрите на специфицирания изборен модел. Данните от втората, по-малка подизвадка (наричана „валидираща“), се използват за изчисляване на изборната вероятност на всяка продуктова алтернатива при всеки респондент, но на базата на параметрите на модела, оценени с данните от оценяващата извадка. Процентът на правилно предсказаните случаи във валидиращата извадка е меродавният критерий, от който се прави изводът за предиктивната способност на модела. При кръстосаното валидиране процесът на подразделяне на изходната извадка на подизвадки с различни съотношения се повтаря многократно и като база за сравнение и избор на „най-добрия“ модел се използват осреднените резултати от всяка итерация. Разбира се, като база за сравняване е възможно да се използва коефициентът λ_{LR} или индексът ρ^2на правдоподобност на двата модела, оценени съответно с данни от оценяващата и валидиращата подизвадки (Norwood et al., 2001).

4. Прогностичен анализ с модели на дискретния избор

В предходната точка бяха синтезирани някои подходи, процедури и критерии, приложими за валидиране на спецификацията на моделите на избора. Ако с тяхна помощ се докаже, че „качествата“ на съставения модел са удовлетворителни, може да се пристъпи към използване на модела за прогностичен анализ.

С понятието прогностичен (или алтернативното „предиктивен“) анализ отбелязваме аналитичен процес на предсказване на изборното поведение на потребителите с помощта на емпирични данни от заявени предпочитания. Прогностичният анализ е най-важният етап от процеса на изучаване на потребителското поведение.

Въпреки че мултиномиалният логит модел се оценява на индивидуално равнище, на практика предсказването на изборното поведение на отделния потребител не е от първостепенно значение. За планирането на маркетингови стратегии и за оптимизиране на маркетинговия микс е от по-голяма важност агрегирането на прогнозната информацията до обобщени индикатори за търсенето (напр. пазарен дял за всяка продуктова алтернатива, ценова еластичност, готовност за плащане), които да са приложими за целия пазар.

4.1. Оценяване на пазарния дял

Оценяването на пазарния дял предполага разработването на модел за предсказване на дела на предпочитанията на всяка алтернатива, изведен чрез агрегиране на индивидуалните вероятности за избор на всяка наблюдавана алтернатива i при всеки респондент n, при наличие на информация за общи, специфични за алтернативите и социодемографски независими променливи x_n.

Процесът на агрегиране на индивидуалните изборни решения до индикатора „пазарен дял“ може да се операционализира по следния начин. Вече обозначавахме нееднократно, че вероятността отделният потребител n (респондент от извадката) да избере продукта i, при наличието на данни за x_n и оценки θ на параметрите на специфициран логит модел, може да се представи като:

$$P_n(i|x_n;\theta).$$

Обръщаме внимание, че в горния израз отсъстват изборните множества C_n, от които респондентът избира желаната алтернатива i, като същевременно припомняме, че изборните множества биха могли да варират при отделните потребители. Причината за отсъствието на информация за C_n се обяснява с обстоятелството, че целта на агрегирането е да се изчисли броят на индивидите в генералната съвкупност $N_T(i)$, които биха избрали конкретната алтернатива i:

$$N_T(i) = \sum_{n=1}^{N_T} P_n(i|x_n; \theta).$$

В общия случай, по-голям интерес за анализиращия представлява относителният дял, а не абсолютният брой потенциални потребители на i. Относителният дял на дадена продуктова алтернатива i може да се интерпретира като дял на предпочитанията, а с известна условност и като пазарен дял. Ако отбележим пазарния дял на продуктовата алтернатива с $MS(i)$, то той би могъл да се изрази като частно от $N_T(i)$ и общия брой потребители, съставящи целевия пазар (респ. размера на извадката N). С други думи пазарният дял представлява очакваната стойност $E[.]$ на осреднената изборна вероятност в генералната съвкупност:

$$MS(i) = \frac{1}{N}\sum_{n=1}^{N_T} P_n(i|x_n; \theta) = E[P(i|x_n; \theta].$$

Горният израз би могъл да се представи в структуриран вид по следния начин (Таблица 4). Всеки ред от таблицата представлява отделен потребител n от популацията N. Всяка колона показва една от продуктовите алтернативи в изборното множество j. Във всяка клетка на таблицата е дефиниран отделен модел на избора, който предсказва вероятността P, с която съответният потребител би избрал кореспондиращата алтернатива.

Таблица 4

Агрегиране на индивидуалните изборни вероятности в модел на пазарния дял

Потребители	Продуктови алтернативи				Общо
$(i = 1, \dots, N)$	1	2	$\cdots$	J	
1	$P(1\|x_1; \theta)$	$P(2\|x_1; \theta)$	$\cdots$	$P(J\|x_1; \theta)$	1
2	$P(1\|x_2; \theta)$	$P(2\|x_2; \theta)$	$\cdots$	$P(J\|x_2; \theta)$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$P(1\|x_1; \theta)$	$P(1\|x_1; \theta)$	$\cdots$	$P(J\|x_N; \theta)$	1
Общо	$N(1)$	$N(2)$	$\cdots$	$N(J)$	N

Източник: (Bierlaire & Lurkin, 2020: 7.1.1 Aggregation).

Резонно е, сумата от изборните вероятности на всички продуктови алтернативи да е единица. Сумата по колони по същество представлява аналитичния израз за очаквания брой потребители, които биха избрали съответна алтернатива j. Така че агрегирането е свързано с обследване на таблицата и изчисляване на суми по редовете. Обикновено тази таблица е твърде голяма. Тя може да има много редове, тъй като в повечето случаи популацията, съставяща целевия пазар, е много голяма. Този проблем би могъл да се реши чрез прилагане на техника на стратифициран подбор. Ако са налице голям брой продуктови алтернативи, таблицата би могла да има и твърде много колони. В този случай за агрегираното й изучаване се прилага техника, известна като микросимулация.

4.1.1. Агрегиране чрез стратифициран подбор

Стратифицираният подбор се състои в подразделяне на генералната съвкупност (респ. на извадката) на G броя взаимноизключващи се и относително еднородни групи (страти). Подобно априорно сегментиране може да се обоснове от логистиката при набирането на данните и целите на проучването. Например всяка страта може да представлява географска територия (напр. населено място, община или област), където да е била провеждана полевата работа по набиране на емпиричните данни. Страти могат да се формират и на базата на социодемографски признаци, като напр. възраст, ако допускаме, че възрастта влияе върху изборното поведение.

След като стратите са предефинирани, от всяка една от тях посредством прост случаен подбор се извличат S_g наблюдения. Общият размер на извадката, съставена от представителите на всички страти, е $S = \sum_{g=1}^{G} S_g$. Тъй като стратите обикновено са с различна големина, за всяка от тях се определя тегло ω_g:

$$\omega_g = \frac{N_g}{N}\frac{S}{S_g} = \frac{\text{Дял на потребителите от сегменг } g \text{ в генералната съвкупност}}{\text{Дял на потребителите от сегмент } g \text{ в извадката}}.$$

За да бъде отчетено обстоятелството, че всеки потребител n принадлежи към точно една страта g, въвеждаме една помощна индикаторна променлива δ_{ng}, която да получава стойност 1, ако $n \in g$, и респ. 0, ако $n \notin g$. Така за всеки респондент n ще е в сила:

$$\omega_g = \sum_{g=1}^{G} \delta_{ng}\omega_g.$$

Оттук извеждането на прогностична оценка на пазарния дял $\widehat{MS}(i)$ на съответната продуктова алтернатива i, би могло да се извърши по формулата:

$$\widehat{MS}(i) = \frac{1}{S}\sum_{n=1}^{S} \omega_n P(i|x_n;\theta).$$

4.1.2. Агрегиране чрез микросимулация

Идеята на микросимулацията се изразява в схващането, че е грешно да се допусне, че всеки път всеки потребител би избрал само и единствено алтернативата с най-висока изборна вероятност (т.е. с най-висока възприемана полезност). Затова, за да се имитира случаен процес, за всеки респондент n е възможно да се генерират R случайни подбора сред алтернативите, отчитайки тяхната оценена вероятност. След всеки подбор се запазва резултатът под формата на дихотомна променлива $\hat{y}_{inr}$, която получава значение 1, ако лицето n се случи да е избрало алтернативата i, и 0 във всички останали случаи. Чрез тези симулирани избори ние всъщност можем да се доближим до изборната вероятност:

$$P_n(i|x_n;\theta) \approx \frac{1}{R}\sum_{i=1}^{R} \hat{y}_{inr}.$$

Вероятността ще бъде просто средният брой пъти, когато е избрана тази алтернатива. Очевидно е, че при $R \to \infty$, симулацията ще възпроизведе действителната структура на вероятностите, изчислена от изборния модел. Така чрез симулираните изборни итерации, се отчита фактът, че всяка алтернатива има ненулева вероятност да бъде избрана.

От резултатите от микросимулацията лесно може да се изведе оценка на пазарния дял на всяка алтернатива i. Просто най-напред е необходимо да се „преброят" респондентите, които биха избрали тази алтернатива:

$$\widehat{N}(i) = \frac{1}{R}\sum_{n=1}^{N}\sum_{i=1}^{R}\hat{y}_{inr}.$$

С тази информация вече е тривиално да се намери относителният дял на хората (т.е. пазарният дял $MS(i)$ от популацията, които биха избрали i:

$$MS(i) = \frac{N(i)}{N} = \frac{1}{N}\frac{1}{R}\sum_{n=1}^{N}\sum_{i=1}^{R}\hat{y}_{inr}.$$

4.2. Оценяване на готовността за плащане

Ако съставеният модел съдържа променливи с данни за цената или стойността на съответните продуктови алтернативи, е възможно да се проведе задълбочен анализ на монетарната оценка за всяка една атрибутивна независима променлива и да се оцени, каква цена би бил готов да плати всеки отделен потребител за всяка една възможна модификация на значенията на всички останали неценови независими променливи. Типичен пример при изучаване на избора на транспортни услуги е цената на времето, т.е. допълнителната цена, която пътникът е готов да плати, за да намали времето за пътуване. В продуктовите изследвания това би могло да бъде допълнителната сума пари, които купувачът е готов да даде, ако предлаганата продуктова алтернатива е с друг цвят, размер или марка.

За да изведем аналитично готовността за плащане на базата на вече оценен изборен модел, въвеждаме следните обозначения:

- с p_{in} бележим цената на алтернатива i, за потребител n;
- с x_{in} отбелязваме стойността на друга, неценова променлива от модела, влияеща върху вероятността, потребител n да избере алтернатива i;
- с $V_{in}(p_{in}, x_{in})$ изразяваме стойността на функцията на полезност, която потребител n асоциира с алтернатива i. В случая под възприемана полезност разбираме субективно възприеманото съотношение между ползата и разходите за потребителя;
- с δ_{in}^{x} бележим възможно изменение (в посока възприемано подобрение) на стойността на променливата x_{in}.
- с δ_{in}^{p} изразяваме сумата пари, които потребителят е готов да заплати допълнително, за да получи същата възприемана полезност.

От последните две обозначения можем да изведем следната релация:

$$V_{in}\left(c_{in}+\delta_{in}^{p},x_{in}+\delta_{in}^{x}\right)=V_{in}(c_{in},x_{in}).$$

Готовността за плащане представлява изменението на разходите за потребителя n, което компенсира изменението на x с единица, т.е. $\delta_{in}^{c}/\delta_{in}^{x}$.

За изчисляването на готовността за плащане е от значение равнището на скалиране на променливата x_{in}. Ако допуснем, че x_{in} е метрична променлива и ако функцията V_{in} е диференцируема спрямо x_{in} и p_{in}, то прилагайки теоремата на Тейлър[6] би следвало, че

$$\frac{\delta_{in}^{p}}{\delta_{in}^{x}}=-\frac{(\partial V_{in}/\partial x_{in})(p_{in},x_{in})}{(\partial V_{in}/\partial p_{in})(p_{in},x_{in})}.$$

Ако x_{in} и p_{in} са в линейна зависимост във функцията на полезността, V_{in}

$$V_{in}(p_{in},x_{in})=\beta_{p}c_{in}+\beta_{x}x_{in}+\cdots,$$

то готовността за плащане ще представлява съотношението между коефициентите пред съответните независими променливи:

$$\frac{\delta_{in}^{p}}{\delta_{in}^{x}}=-\frac{\left(\frac{\partial V_{in}}{\partial x_{in}}\right)(p_{in},x_{in})}{\left(\frac{\partial V_{in}}{\partial p_{in}}\right)(p_{in},x_{in})}=-\frac{\beta_{x}}{\beta_{c}}.$$

Последният израз показва начина на изчисляване на готовността за плащане при нарастване на стойността на променливата x_{in}. Ако това нарастване води след себе си увеличаване на полезността на алтернативата i, то би трябвало $\beta_{x}>0$. Тъй като е логично, параметърът пред цената да е отрицателно число ($\beta_{p}<0$), то готовността за плащане ще е положителна монетарна оценка. Готовността за плащане би могла да бъде отрицателна, ако намаляването на цената компенсира някакво изменение в x_{in}, което провокира намаляване на полезността. Понякога този „феномен“ се нарича готовност за приемане на обезщетение (Martín-Fernndez et al., 2010).

4.3. Оценяване на еластичността на търсенето

Друг „стандартен“ индикатор за анализ на търсенето, който може да се изведе на базата на оценен изборен модел, е еластичност. Еластичността в най-широк смисъл изразява способността на зависимата променлива да реагира на изменения на независимите променливи в даден иконометричен модел. В контекста на изборните модели това означава да оценим относителното изменение на изборната вероятност вследствие на едно относително изменение на някоя от независимите променливи на модела (Lovell, 2004, pp. 75–79).

Обикновено се прави разлика между директна и кръстосана еластичност. Директната еластичност отразява относителното изменение на изборната вероятност на алтернатива i, породено от изменение на някоя от променливите на същата алтернатива. Изразява се като:

[6] Вж. Теорема на Тейлър (н. а.), категория „Математически анализ“ на Уикипедия, последен достъп на 20.11.2020 г. от https://bit.ly/2I5BNbx.

$$E_{x_{ink}}^{P_n(i)} = \frac{\Delta P_n(i)}{\Delta x_{ink}} \frac{x_{ink}}{P_n(i)},$$

където с Δx_{ink} бележим абсолютното изменение на променливата x_{ink}.

Кръстосаната еластичност отразява относителното изменение на изборната вероятност на алтернатива i, породено от изменения на някоя от променливите на друга алтернатива от изборното множество. Ако с Δx_{jnk} отбележим абсолютното изменението на променливата x_{jnk}, то моделът за изчисляване е:

$$E_{x_{jnk}}^{P_n(i)} = \frac{\Delta P_n(i)}{\Delta x_{jnk}} \frac{x_{jnk}}{P_n(i)}.$$

Двете дефиниции описват т.нар. „дъгова“ или „средна“ еластичност, тъй като изразяват еластичността в определен диапазон на функцията. Базисните значения $\frac{x_{ink}}{P_n(i)}$ и респ. $\frac{x_{jnk}}{P_n(i)}$ на променливите се определят като средна величина от равнищата им в началото и края на диапазона на изменение. Средната еластичност е на практика най-полезна за иконометричен анализ. Ако обаче независимите променливи са непрекъснати и ако измененията им клонят към безкрайно малко число, т.е. $\Delta x_{ink} \to \infty$, може да се изведе и „точкова“ еластичност.

Представените форми на директна и кръстосана еластичност са по същество дезагрегирани. Това означава, че с посочените формули е възможно изчисляване на еластичността на индивидуално равнище (т.е. на равнище потребител n). Интерес би представлявало обаче да се изведе и агрегирано оценяване на еластичността на конкретна продуктова алтернатива i, но на нива пазари или пазарен сегмент. За целта нека да изходим от познатия ни вече агрегиран модел на пазарния дял, изведен на базата на индивидуалните изборни вероятности и стратифициран подбор (Bierlaire & Lurkin, 2020: 7.3 Indicators):

$$\widehat{MS}(i) = \frac{1}{S} \sum_{n=1}^{S} \omega_n P(i|x_n;\theta).$$

Приемаме, че независимите променливи x_n са метрични непрекъснати и че относителните изменения на тези променливи са еднакви при всеки индивид от генералната съвкупност (респ. респондент от извадката), т.е.:

$$\frac{\Delta x_{ink}}{x_{ink}} = \frac{\Delta x_{ipk}}{x_{ipk}} = \frac{\Delta x_{ik}}{x_{ik}},$$

като $x_{ik} = \frac{1}{S}\sum_{n=1}^{S} x_{ink}$, а n и p са индексите на два различни респондента от извадката. Изхождайки от тези презумпции, агрегираната директна дъгова еластичност по отношение на средната стойност на x_{ik}

$$E_{x_{ik}}^{\widehat{W}(i)} = \frac{\Delta \widehat{W}(i)}{\Delta x_{ik}} \frac{x_{ik}}{\widehat{W}(i)}.$$

Изразявайки $\Delta\widehat{MS}(i)$ вследствие на Δx_{ik} от агрегирания модел на пазарния дял и замествайки в горния израз, се получава:

$$E_{x_{ik}}^{\widehat{MS}(i)} = \frac{1}{S} \sum_{n=1}^{S} \omega_n \frac{\Delta P_n(i|x_n, C_n)}{\Delta x_{ik}} \frac{x_{ik}}{\widehat{MS}(i)}.$$

Тъй като $\frac{\Delta x_{ik}}{x_{ik}} = \frac{\Delta x_{ink}}{x_{ink}}$, изразът може да се изпише и като:

$$E_{x_{ik}}^{\widehat{MS}(i)} = \frac{1}{S}\sum_{n=1}^{S} \omega_n \frac{\Delta P_n(i|x_n,C_n)}{\Delta x_{ik}} \frac{x_{ink}}{\widehat{MS}(i)}.$$

Ако умножим дясната страна на последното уравнение с $\frac{P_n(i|x_n,C_n)}{P_n(i|x_n,C_n)}$ и след съответното преобразуване, един от множителите показва дефиниция за дезагрегирана средна еластичност:

$$E_{x_{ik}}^{\widehat{MS}(i)} = \frac{1}{S}\sum_{n=1}^{S} \omega_n \frac{\Delta P_n(i|x_n, C_n)}{\Delta x_{ik}} \frac{x_{ink}}{P_n(i|x_n, C_n)} \frac{P_n(i|x_n, C_n)}{\widehat{MS}(i)} =$$

$$\frac{1}{S}\sum_{n=1}^{S} \omega_n E_{x_{ink}}^{P_n(i)} \frac{P_n(i|x_n,C_n)}{\widehat{MS}(i)}.$$

И накрая, замествайки формулата за $\widehat{MS}$(i), се получава окончателният модел за изчисляване на агрегираната средна еластичност на алтернатива i от x_{ink}:

$$E_{x_{ik}}^{\widehat{MS}(i)} = \frac{1}{S}\sum_{n=1}^{S} E_{x_{ink}}^{P_n(i)} \frac{\omega_n P_n(i|x_n,C_n)}{\sum_{n=1}^{S} \omega_n P(i|x_n;C_n)}.$$

От последния израз е видно, че агрегираната еластичност представлява претеглена сума от дезагрегираните еластичности на отделните респонденти, като теглата се явяват нормализирана версия на теглата ω_n, използвани при агрегиране на пазарния дял.

Извеждането на агрегираната кръстосана ценова еластичност следва абсолютно същата логика. Ако я следваме, крайният модел за изчисляване придобива вида:

$$E_{x_{jk}}^{\widehat{MS}(i)} = \frac{1}{S}\sum_{n=1}^{S} E_{x_{jnk}}^{P_n(i)} \frac{\omega_n P_n(i|x_n,C_n)}{\sum_{n=1}^{S} \omega_n P(i|x_n;C_n)}.$$

Заключение

В настоящата студия бяха разгледани ограничен набор от базисни аналитични модели и техники за анализ на дискретния избор, които имат съществено значение за анализ на изборното поведение и вземането на стратегически и тактически маркетингови решения. Посочени бяха някои насоки за тяхното оценяване и използване. Въз основа на общия модел на дискретния избор е възможно да бъдат извеждани и други обобщени и/или дезагрегирани индикатори за анализ на търсенето, като например за оценяване на потребителския излишък, представляващ разликата между готовността за плащане и действително платената цена (McConnell, 1995). Освен това той може да бъде прилаган като основа при съставянето на иконометрични модели за управление и оптимизиране на приходите (Talluri & Van Ryzin, 2004), както и за съставянето на симулационни модели за условен сценариен анализ (Orme & Chrzan, 2017, p. 199).

Извън предмета на настоящото изследване остават проблемите, свързани със събирането на изборни данни, планирането на изборни експерименти, специфициране на йерархични, смесени и вложени логит модели, както и модели с подреден избор. Встрани от вниманието и обект на последващо изследване остават и някои важни алтернативни числови алгоритми за оценяване параметрите на моделите (като например пробит или йерархична бейсовска регресия). Всички описани методи и процедури изискват задълбочени иконометрични знания, но е възможно да се реализират сравнително лесно със софтуерни пакети с отворен код. На първо място препоръчваме базирания на програмния език Python пакет Biogeme (Bierlaire, 2020), както и базирания на програмния език R Apollo (Hess & Palma, 2019). Като алтернативи могат да се ползват и LARCH (Newman, 2016), R's mlogit library (Croissant, 2020) и R's RSGHB library. Съществуват и множество комерсиално предлагани софтуерни програми, позволяващи анализирането на данни от заявен или разкрит избор и оценяването на разнообразни изборни модели. По-известните са например NLOGIT (Econometric Software, Inc.) STATA, Discrete Choice Bundle (StatWizards), JMP, Q Research Software, CBC/HB System (Sawtooth Software) , Latent GOLD-Choice (Statistical Innovations), както и комерсиално достъпните софтуерни среди за числов анализ MATLAB (MathWorks), GAUSS (Aptech).

Използвани източници

Alves, I. F., & Neves, C. (2011). Extreme Value Distributions. In *International Encyclopedia of Statistical Science*. https://doi.org/10.1007/978-3-642-04898-2_246

Axhausen, K. W., Hess, S., König, A., Abay, G., Bates, J. J., & Bierlaire, M. (2008). Income and distance elasticities of values of travel time savings: New Swiss results. *Transport Policy*. https://doi.org/10.1016/j.tranpol.2008.02.001

Ben-Akiva, M., & Lerman, S. R. (1985). Tests and Practical Issues in Developing Discrete Choice Models. In *Discrete Choice Analysis: Theory and Application to Travel Demand*.

Bierlaire, M. (2015). *Optimization: Principles and Algorithms*. EPFL Press.

Bierlaire, M. (2020). A Short Introduction to PandasBiogeme. In *Report TRANSP-OR 200605*. https://transp-or.epfl.ch/documents/technicalReports/Bier20.pdf

Bierlaire, M., & Lurkin, V. (2017). Introduction to Disaggregate Demand Models. In *The Operations Research Revolution*. https://doi.org/10.1287/educ.2017.0169

Bierlaire, M., & Lurkin, V. (2020). Introduction to Discrete Choice Models. In *MOOC lecture (edX)*.

Carson, R. T., & Czajkowski, M. (2014). The discrete choice experiment approach to environmental contingent valuation. In *Handbook of Choice Modelling*. https://doi.org/10.4337/9781781003152.00015

Chandukala, S. R., Kim, J., Otter, T., Rossi, P. E., & Allenby, G. M. (2007). Choice models in marketing: Economic assumptions, challenges and trends. *Foundations and Trends in Marketing*, *2*(2), 97–184. https://doi.org/10.1561/1700000008

Cox, D. R. (1962). Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society: Series B (Methodological)*. https://doi.org/10.1111/j.2517-6161.1962.tb00468.x

Cox, David R. (1961). Tests of separate families of hypothesis. *Proc. 4th Berkeley Symp. on Math. Stat. and Probability*.

Croissant, Y. (2020). Package 'mlogit.' In *Discrete Choice Methods with Simulation, Second Edition*.

Davidson, R., & MacKinnon, J. G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica*. https://doi.org/10.2307/1911522

Desarbo, W. S., Ansari, A., Chintagunta, P., Himmelberg, C., Jedidi, K., Johnson, R., Kamakura, W., Lenk, P., Srinivasan, K., & Wedel, M. (1997). Representing heterogeneity in consumer response models: 1996 Choice conference participants. *Marketing Letters*. https://doi.org/10.1023/A:1007916714911

Fisher, G., & McAleer, M. (1979). On the interpretation of the cox test in econometrics. *Economics Letters*. https://doi.org/10.1016/0165-1765(79)90225-8

Gumbel, E. J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications*.

Hess, S., & Palma, D. (2019). Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*. https://doi.org/10.1016/j.jocm.2019.100170

Kamakura, W. A., Kim, B. Do, & Lee, J. (1996). Modeling Preference and Structural Heterogeneity in Consumer Choice. *Marketing Science*. https://doi.org/10.1287/mksc.15.2.152

King, G. (2010). Unifying political methodology: The likelihood theory of statistical inference. In *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. https://doi.org/10.2307/1963270

Koppen, M. (2001). Characterization Theorems in Random Utility Theory. In *International Encyclopedia of the Social & Behavioral Sciences*. https://doi.org/10.1016/b0-08-043076-7/00600-8

Larsen, R. J., & Marx, M. L. (2018). *An introduction to mathematical statistics and its applications* (Sixth). Pearson.

Louviere, J., & Eagle, T. (2006). Confound it! That pesky little scale constant messes up our convenient assumptions. *Proceedings of the Sawtooth Software Conference.*, 211–228.

Lovell, M. C. (2004). Economics with calculus. In *Economics with Calculus*. https://doi.org/10.1142/5523

Luce, R. D. (2014). Individual choice behavior: A theoretical analysis. In *Individual choice behavior: A theoretical analysis.* https://doi.org/10.1037/14396-000

Marley, A. A. J. (1968). Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*. https://doi.org/10.1016/0022-2496(68)90078-3

Martín-Fernndez, J., Del Cura-Gonzlez, M. I., Gámez-Gascán, T., Oliva-Moreno, J., Domínguez-Bidagor, J., Beamud-Lagos, M., & Pérez-Rivas, F. J. (2010). Differences between willingness to pay and willingness to accept for visits by a family physician: A contingent valuation study. *BMC Public Health*. https://doi.org/10.1186/1471-2458-10-236

McConnell, K. E. (1995). Consumer surplus from discrete choice models. *Journal of Environmental Economics and Management*. https://doi.org/10.1006/jeem.1995.1046

McFadden, D. (1980). Econometric Models for Probabilistic Choice Among Products. *The Journal of Business*. https://doi.org/10.1086/296093

McFadden, D. L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*.

Nadarajah, S., & Kotz, S. (2005). A generalized logistic distribution. *International Journal of Mathematics and Mathematical Sciences*. https://doi.org/10.1155/IJMMS.2005.3169

Newman, J. (2016). LARCH: A package for estimating multinomial, nested, and cross-nested logit models that account for semi-aggregate data. *56th Annual AGIFORS Symposium*. https://larch.newman.me/

Nicholson, W., & Snyder, C. (2017). Microeconomic theory: basic principles and extensions. In *Microeconomic Theory: Basic Principles & Extensions* (12th ed.). Cengage Learning.

Norwood, B., Ferrier, P., & Lusk, J. (2001). Model selection criteria using likelihood functions and out-of-sample performance. *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*.

Orme, B. K., & Chrzan, K. (2017). *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros*. Sawtooth Software, Inc.

Orme, B. K., & Johnson, R. M. (2007). A New Approach to Adaptive CBC. *Sawtooth Software Research Paper Series*.

Paczkowski, W. R. (2016). *Market Data Analysis Using JMP®*. SAS Institute Inc.

Swait, J., & Louviere, J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research*. https://doi.org/10.2307/3172883

Talluri, K., & Van Ryzin, G. (2004). Revenue Management Under a General Discrete Choice Model of Consumer Behavior. In *Management Science*. https://doi.org/10.1287/mnsc.1030.0147

Train, K. E. (2009). Discrete choice methods with simulation, second edition. In *Discrete Choice Methods with Simulation, Second Edition* (Second). https://doi.org/10.1017/CBO9780511805271

Whitten, S. (2004). Designing the Choice Modelling Survey Instrument for Establishing Riparian Buffers in the Fitzroy Basin. In *Establishing the potential for offset trading in the lower Fitzroy River Research Repoert*.