joforum74 2008. 03. 18. 00:15 Page 43

Mikroökonometriai modellek alkalmazása a marketingkutatásban – a Tobit regressziómodell

LÁZÁR EDE

A számítástechnika gyors fejlődése és elterjedése nemcsak a vállalatok üzleti folyamataira gyakorolt stratégiai jelentőségű hatást, hanem a gyorsabb és komplexebb üzletiinformáció-elemzési igények visszahatottak, generálták az elméleti modellek fejlődését is. A különböző kvantitatív (ezen belül az ökonometriai) modellek üzleti alkalmazásainak robbanásszerű fejlődéséhez vezetett, hogy a nagyon számításigényes módszerek kipróbálása már nem csak egy-két akadémiai fehérköpenyes kiváltsága volt, akik hozzáférhettek a nagygépes rendszerekhez, hanem „bármely" egyetem vagy vállalat elemzője előtt megnyílt a lehetőség modellek fejlesztésére és kipróbálására a valós üzleti adatokon.

Az 1950-es évek menedzsmentirodalmában és -gyakorlatában jelentős szerepet betöltő kvantitatív módszerekkel, főképp operációkutatási modellekkel kapcsolatos várakozások és szép remények a 60-as évekre jórészt szertefoszlottak, egyértelművé váltak e modellek alkalmazásainak korlátai (lásd pl. Herbert Simon munkáit). A számítási kapacitás gyors növekedése és az elemzők-fejlesztők táborának kiterjedése az akadémiai szféráról a kis- és középvállalati szintig odavezetett, hogy a 90-es évektől a kvantitatív módszerek, ezen belül a többváltozós statisztikai, ökonometriai modellek nemcsak a makroökonómiai elemzések eszköztára maradt, hanem a mindennapi üzleti döntések támogatásának eszközévé vált. Ez a fejlődés nyilván az akadémiai elit figyelmét sem kerülte el; a 2000. évi közgazdasági Nobel-díjat James Heckman és Daniel McFadden a mikroökonómiai szereplők (vállalat, háztartás, fogyasztó) empirikus elemző eszköztárának fejlesztéséért, ezen belül főképp a kvalitatív és korlátozott eredményváltozójú modellek kidolgozásáért kapták. A régi-új, alkalmazott tudományterület – amelyet egyre többen *mikroökonometriának* neveznek[1] – megjelent a vezető egyetemek tanrendjében.

A Tobit-modell

Empirikus kutatások nagyon gyakori problémája az adatok *korlátozott* jellege. Egy változót korlátozottnak nevezünk, ha a megfigyelt értékek folytonosak egy bizonyos terjedelemben, majd egy bizonyos küszöbértéknél (gyakran nullánál) tartalmaznak egy nem lineáris ugrást. Példaként tekintsünk egy igencsak valószerű marketingkutatási kérdést: „Mennyit fizetne Ön ezért a termékért/szolgáltatásért?". A válaszok két csoportba sorolhatók: azok, akik nem kívánják megvenni a terméket/szolgáltatást, nullát, a többiek pedig egy többé-kevésbé széles skálán szóródó értéket neveznek meg.

[1] Jerry Hausman: Microeconometrics. *Journal of Econometrics*, Volume 100, Issue 1, 2001/1. 33–35.

joforum74 2008. 03. 18. 00:15 Page 44

Long (1997) szerint amíg a korlátozott eredményváltozójú regressziós modelleket viszonylag régóta alkalmazzák a biometrikai, mérnöki kutatásokban, az első közgazdasági alkalmazás James Tobin 1958-as cikkében van megjelentetve. A későbbi Nobel-díjas közgazdász – aki főképp makro-pénzügyi eredményei alapján ismert – a háztartások tartós fogyasztási cikkekre, történetesen személyautóra fordított kiadásait vizsgálta ezzel a modellel. Ez a cikk nagy hatást fejtett ki a hetvenes évek elméleti és empirikus kutatásaira, a szerteágazó modellfejlesztéseket többen „Tobin probit" modellnek, vagy *„tobit regressziónak"* nevezték. Ma már az empirikus kutatások irodalmában megtalálható példák az alkalmazások széles körét ölelik fel: a nők munkapiacon eltöltött munkaóráinak a száma (Quester–Greene 1982), bankok hatékonysága (Stavarek 2003), piaci volatilitás és jövedelemegyenlőtlenségek (Huck–Normann–Oechssler 2000), európai felsőoktatási intézmények relatív hatékonysága (Joumady–Ris 2002), az USA külföldi beruházásainak (FDI) hatása a fejlődő országok munkapiacára (Lipsey 2004). Az elmúlt egy-két évben a modell oktatása megjelent néhány magyarországi és romániai közgazdasági mesteri és doktori képzési programban is.

De miért is nem használhatjuk a korlátozott függő változót tartalmazó keresletelemzéseknél a jól bevált lineári regressziós modellt, a konzisztens becslést adó OLS-sel? A következő gyakran előforduló, de téves megoldásokkal találkozhatunk a korlátozott adatokból származó függő változók regressziós modellezése kapcsán. Tekintsük Tobin személyautó-vásárlás kutatására irányuló példáját.

1. Ha lineáris regresszió segítségével modellezzük, hogy milyen tényezők határozzák meg az autóvásárlásra fordított összeget (Y függő változó), és a minta valamennyi esetét bevonjuk az elemzésbe, akkor a becsült regressziós függvénynek értelmezhetetlen negatív értékei is lesznek, és ez a függvény (nagyon) rossz illeszkedését eredményezi.

2. További lehetőség, hogy egy *logisztikus regressziós* modellt alkalmazzunk Y=0 (nem vásárol) versus az Y=1 (vásárol) valószínűségek becslésére. Ebben az esetben egy jó becslést kaphatunk arra, hogy kik vásárolnak autót (milyen tényezők befolyásolják a vásárlás valószínűségét), de elveszítjük azt az információt, hogy mennyit hajlandóak költeni.

3. Végül a talán leggyakoribb félmegoldás: lazán mellőzzük a mintánkból a nem vásárlókat, és egy lineáris regressziós modellt specifikálunk az Y>0 értékekre. Ez a modell jelentős torzítottsághoz és a függő, illetve a független változók kapcsolatának alulbecsléséhez vezet.

A Tobit-modell értelmezhető úgy, mint egy lineáris regresszió, amelyben a függő változót vagy jobbról, vagy balról, vagy mindkét irányból valamilyen küszöbérték korlátozza. A modell általános formája Greene alapján (2003) két küszöbérték esetén:

joforum74 2008. 03. 18. 00:15 Page 45

$$y^*_i = \alpha + \beta x_i + \varepsilon, \quad (1)$$
$$y_i = a \text{ ha } y^*_i =< a,$$
$$y_i = b \text{ ha } y^*_i >= b,$$
$$y_i = y^*_i \text{ különben,}$$

ahol y^*_i ún. látens változó, amelyre az esetek egy részében (a küszöbértékeken kívül) nincs megfigyelésünk, *y* a megfigyelt függő változó, *a* és *b* pedig a regressziós paraméterek.

A legkisebb négyzetek módszere (OLS) nem alkalmas eljárás a regressziós paraméterek becslésére, ezért az alkalmazott módszer a maximum likelihood (MLE). Az OLS-sel szemben, amely a becsült és a megfigyelt értékek közötti távolságot minimalizálja, az MLE a log likelihood függvény maximalizálásán alapul. A log likelihood a likelihood függvény logaritmusa, amely egy valószínűséget fejez ki. A zéró küszöbértékű modell likelihood függvénye két részből áll: $L=P_1P_2$

$$P_1 = \prod_{i=1,\,m} 1/\sigma * f[(Y_i - \alpha - \beta X_i)/\sigma] \quad \text{ha } Y>0 \quad (2)$$

$$P_2 = \prod_{j=1,\,n} F[(-\alpha - \beta X_j)/\sigma] \quad \text{ha } Y=0,$$

ahol *f* sűrűségfüggvény, *F* eloszlásfüggvény, *m* azon részmintaelemek száma, ahol Y pozitív, és *n* az Y=0 esetek száma.

A modell építésének folyamata hasonló a lineáris regressziós modellekéhez, iteratív lépéseken keresztül vonjuk be a magyarázó változókat és teszteljük a hatásukat. Egyértelmű hátránya a modellnek, hogy feltételezi az eltérésváltozó (ε) normál eloszlását, heteroszkedaszticitás esetén a módszer a paraméterek torzított becsléséhez vezet.

Marginális hatás

Általában a többváltozós regressziós modellek nagy erőssége, hogy a függő változóra szignifikánsan ható tényezők (független változók) meghatározása mellett megbecsülhetjük a hatás intenzitását is. Amíg a lineáris regressziós egyenletben egy x_i magyarázó változó – a többi magyarázó változó rögzített értéke melletti – ceteris paribus hatása a függő változóra egyenlő annak paraméterével, a Tobit-modellben lényegesen bonyolultabb ennek kiszámítása. Tekintsük adottnak a fenti általános képletet y^* látens és *y* megfigyelt függő változóval, *a* és *b* konstansokkal. Feltételezzük továbbá, hogy az ε hibatag folytonos véletlen változó, zéró átlaggal és σ^2 állandó varianciával, azaz $f(\varepsilon|x) = f(\varepsilon)$. Ez esetben a határhatás képlete:

$$\delta E(y_i|x_i) / \delta x_i = (F_b - F_a) * \beta = \beta * \text{Prob}[a < y_i^* < b]. \quad (3)$$

joforum74 2008. 03. 18. 00:15 Page 46

Bizonyítása és részletes tárgyalása megtalálható Greene (2003) könyvében. Ez az általános forma két küszöbérték esetére érvényes. Abban a gyakori esetben, amikor a küszöbérték zéró és feltételezzük a hibatag normális eloszlását, a következő képlettel számoljuk a határhatást:

$$\delta E(yi|xi) / \delta xi = \beta * Prob(yi^* > 0|x) = \beta *[1 - F(-\beta * xi / \sigma)]. \quad (4)$$

A Tobit-modellben tehát – akárcsak más nem lineáris regressziós modelleknél – egy változó marginális hatása nemcsak az adott változó paraméterbecslésétől függ, hanem a többi független változó értékétől és paraméterbecslésétől is.

Árrugalmassági együttható mint marginális hatás

A marginális hatás viszonylag bonyolult kiszámítása ellenére kiemelten fontos szerepet kap a marketingkutatásban egy új termék beárazását célzó ártesztnél. Amennyiben a modell tartalmazza a termék/szolgáltatás árát mint független változót (egy ártesztnél meglepő lenne, ha nem), akkor az árrugalmassági együtthatót az ár határhatásából képezhetjük. Az árrugalmasság jól ismert általános formája:

$$\varepsilon_p = \delta Q/\delta p * p/Q, \quad (5)$$

természetesen *p* ebben az egyenletben a szolgáltatás árát és nem a vásárlás valószínűségét jelenti, *Q* pedig a keresleti függvényt, ami megegyezik az y* függő változóval. A vásárlás valószínűségét kifejező polinom függvény megfeleltethető a keresleti függvénynek, így a képlet első tagja ($\delta Q/\delta p$) az ár szerinti határhatást jelenti. Felhasználva tehát a (4), (5), (6) egyenleteket, kutatásunkban az árrugalmassági együttható a következő képlettel fejezhető ki:

$$\varepsilon_p = \delta Q/\delta p * p/Q = \beta(1- F(-\beta * x_i / s) * p/y^*, \quad (6)$$

ahol p az ár változót, *y** pedig a látens függő változót jelöli. Amint a képletből is kiderül (ε_p függ x_i-től, ahol i 1-től n-ig tart), az árrugalmasság valamennyi adatsor, azaz valamennyi mintabeli interjúalany esetében kiszámítódik, egy új változót képezve az adattáblában. Felmerül az árrugalmasság egy értékkel való kifejezésének a problémája. Két lehetőségünk van: a rögzítés és az átlagolás módszere. Kutatásunkban az utóbbit használtuk: az esetekhez tartozó árrugalmassági együtthatók mintabeli átlagával számoltunk.

Figyelembe véve, hogy egy termék értékesítéséből származó árbevétel akkor maximális, ha az árrugalmassági együttható -1, a modellünkben addig kell módosítani az árat mint független változót, amíg az árrugalmassági együttható eléri ezt az értéket.

joforum74 2008. 03. 18. 00:15 Page 47

Empirikus kutatás

A következőkben egy valós marketingkutatási projekten keresztül mutatom be a Tobit-modell alkalmazását a kereslet előrejelzésben, előbb kissé részletezve azt a piaci környezetet, ami meghatározta a marketingkutatási problémát. Az utóbbi évek valószínűleg legfontosabb fejleménye az európai országok telekommunikációs piacain a harmadik generációs mobiltelefon-szolgáltatások (3G) technikai kifejlesztése és piaci bevezetése volt, mint a telekommunikáció, internet és adatátviteli technikák közötti konvergencia egyik fontos állomása. A történet az ezredforduló körül kezdődött, de mára Romániában is a mindennapok része lett, nemcsak a multik nyújtanak ilyen szolgáltatást, hanem a hazai tulajdonú távközlési szolgáltató is most építi ki ezt a jelentős beruházást igénylő rendszert.

Közvetlenül az ezredforduló előtti dotcom-lázban a távközlési cégek sok milliárd dollárt fizettek néhány nyugat- és észak-európai országban a licenszdíjakért anélkül, hogy előtte felmérték volna az említett szolgáltatások várható keresletét. A hatalmas licenszdíjak és a dotcom-lufi kipukkanása, a dotcom-cégek tőzsdei árfolyamzuhanása elvezetett oda, hogy néhány, amúgy gyorsan növekvő piaccal és a dotcom-cégek egy jó részével szemben reális üzleti modellel rendelkező telekommunikációs cég is pénzügyileg ellehetetlenedett, illetve csődbe ment. Ez a magyarázata annak, hogy a letisztult piacon megmaradt távközlési cégek sokkal óvatosabban fogtak a 3G szolgáltatások bevezetésének előkészítéséhez, a távközlésitechnológia-fejlesztésben és a legnagyobb kereslettel rendelkező országokban is csak 2003/2004-ben indultak el az első szolgáltatások. A termékbevezetési folyamatnak, ha a licenszdíjak kifizetéséhez képest késve is, de része lett a keresleti oldal marketingkutatási eszközökkel történő vizsgálata.

A következőkben részletezett empirikus kutatási példában az egyik egyesült királyságbeli mobiltelefon-szolgáltató 2001-es marketingkutatásából származó adatoksegítségével mutatom be a módszertant. Lehetőségem adódott a Taylor Nelson Sofres nemzetközi piackutató cég Telecom csoportjának a tagjaként részt venni ebben a projektben.[2] A kutatás fő célja a szélessávú adatátvitelt lehetővé tevő technológiára épülő szolgáltatások (pl. videoletöltés, e-mail, on-line játékok, e-learning) potenciális keresletének, optimális árának és az ezekből adódó várható legnagyobb árbevételnek becslése. Optimális ár alatt természetesen az árbevétel maximalizálása melletti árat értjük.

Az adatgyűjtés technikája személyes megkérdezés volt, amely során az akkor még újdonságnak számító CAPI-eszközöket[3] használtuk. Az árteszt során a vizsgált 3G

[2] A kutatás elvégzése óta eltelt néhány év, de az adatok továbbra is a megbízó tulajdonát képezik. Ezért a tanulmányban csak annyi adatot mutatok be, amennyit a módszertan megértéséhez szükségesnek tartok, és ami alapján nem rekonstruálható a szolgáltatás beárazására vonatkozó semmilyen üzleti információ.

[3] Computer Aided Personal Interview – a személyes megkérdezést támogató leegyszerűsített laptop tartalmazza a kérdőívet, így nemcsak az adatrögzítés idejét és költségét lehet megspórolni, hanem képekkel, rövidfilmekkel a vizsgált terméket/szolgáltatást is jobban be lehet mutatni.

joforum74 2008. 03. 18. 00:15 Page 48

szolgáltatások közül néhány egyetlen havidíjjal lett tesztelve, mások pedig a havidíj mellett forgalmi díjat is tartalmaztak. Természetesen valamennyi tesztelt szolgáltatás optimális beárazása külön modell felállítását és specifikálását jelentette. A havidíjjal tesztelt szolgáltatások esetében az alkalmazott modell a logisztikus regresszió (binomiális és multinomiális) volt, míg a havidíjjal és forgalmi díjjal egyaránt rendelkező szolgáltatások ártesztjénél a Tobit regressziós modellt használtuk. Ebben a tanulmányban csak a Tobit-modellel foglalkozom.

Az interjúk során használt CAPI-eszközök lehetővé tették, hogy a vizsgált szolgáltatás (pl. videoletöltés) tervezett használatának a mennyisége (perc/nap) alapján azonnal kiszámoljuk a szolgáltatás teljes havi költségét (havidíj + használat∗percdíj), és ezt az interjúalannyal interaktív módon egyeztetve addig módosítsuk a használat mennyiségét, amíg a teljes havi költség elfogadható számára. Az eredmény nemcsak az adatok jó minősége és konzisztenciája lett, hanem egy olyan változó, amely a minta egy részében zéró értéket, a maradék részben pedig egy adott terjedelemben szóródó egész értéket vesz fel. Ez a korlátozott jellegű adatsor alkalmas függő változója lehetett a Tobit regressziós modellnek.

Empirikus eredmények

Az első szakasz a modellspecifikáció, a Tobit regressziós egyenlet paramétereinek a meghatározása.[4] A legjobb magyarázó erejű modell megtalálása meglehetősen hosszadalmas folyamatából csak a végeredményt mutatom be. A használt TSP programnak nincs menürendszere, és a Tobit-modell szintaxisa sem a legegyszerűbbek közül való:

```
TOBIT  dmail c preml inchoho tmnbill grndtot age male ftech segm1;
y0=belu .le. 0;  y1=belu .gt. 0;
frml xb1eq xb1 = z0 + z1*preml + z2*inchoho + z3*tmnbill + z4*grndtot + z5*age
+ z6* ftech + z7* male + z8*segm1;
frml eq1 log1=y0*lcnorm(-xb1/sigma)+y1*(-log(sigma)+lnorm((belu-xb1)/sigma));
eqsub(name=tobit) eq1 xb1eq ;
```

A regressziós egyenlet tehát a vásárlás valószínűségét becsli nyolc független változó alapján.

$$P(Y>0) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_8x_8, \qquad (7)$$

ahol: P(Y>0) – a vásárlás valószínűsége, b_0 – konstans, x1 – a szolgáltatás ára, x2 – az interjúalany havi jövedelme, x3 – a jelenlegi mobiltelefon-számla, x4 – valamennyi

[4] A marketingkutatásban legelterjedtebb SPSS program nem tartalmazza (legalábbis a kutatás idején használt 9-es verzió) a Tobit regressziós modellt, ezért a TSP (Time Series Processor) ökonometriai szoftvert használtuk.

joforum74 2008. 03. 18. 00:15 Page 49

3G-szolgáltatás tervezett havi költsége, x5 – életkor, x6 – a technikai újdonságok iránti nyitottság faktora, x7 – nem (0-nő, 1-férfi), x8 – célszegmensbe tartozás (a szegmentációs klaszter változó egyik kategóriájának a dummy-ja) .

Az output legfontosabb eredményei a maximum likelihood módszerrel becsült együtthatók és a hozzá tartozó t-statisztikák.

Dependent variable: DMAIL
Number of observations = 1323 Schwarz B.I.C. = 1407.109
Number of positive obs. = 353 Log likelihood = -1349.607

Parameter	**Estimate**	**St. Error**	**t-statistic**	**P-value**
C	-6.41208	0.870686	-7.3644	[.000]
PREML	-0.10457	0.040461	-2.58459	[.010]
INCHOHO	2.32743	0.417325	5.577014	[.000]
TMNBILL	0.053346	3.36E-03	15.89977	[.000]
GRNDTOT	1.166837	0.363609	3.209045	[.001]
AGE	-0.04805	0.0153	-3.14037	[.002]
FTECH	2.346855	0.473127	4.960305	[.000]
MALE	0.990342	0.340333	2.909917	[.004]
SEGM1	0.94647	0.361213	2.620252	[.009]
SIGMA	4.369107	0.180491	24.20676	[.000]

Standard Errors computed from analytic second derivatives (Newton)

A független változók együtthatóinak vizsgálata értékes információt szolgáltat a termék bevezetését célzó marketingstratégia számára, megmutatja, hogy mely fogyasztói jellemzők befolyásolják a vásárlás valószínűségét, és alapot nyújthat a piac további szegmentációjához. Minket azonban most csak a termék „beárazása", az optimális ár meghatározása érdekel.

A következőkben az árrugalmassági együttható (e_preml) áttekinthetőbb kiszámításához néhány segédváltozót is definiáltunk (pdf, cdf, yt). A Tobit regressziós modell rögzített együtthatói és egy meghatározott áron (preml=0.25) számítjuk ki az árrugalmasság értékét.

```
const z0 – z8 sigma preml 0.25;                                    (8)
genr xb1eq;
genr pdf=norm(xb1/sigma);
genr cdf=cnorm(xb1/sigma);
genr yt =(cdf*xb1)+(pdf*sigma);
genr e_prel=(cdf*z1)*(preml/yhat);
print e_preml;
```

joforum74 2008. 03. 18. 00:15 Page 50

A továbbiakban már csak néhány iteratív lépés az elemzés: addig módosítjuk modellünkben az árváltozó értékét (a többi változó rögzítése mellett), amíg az árrugalmassági együttható -1 lesz. Az optimális árat visszahelyettesítve a keresleti függvénybe, meghatározható az ehhez az árhoz tartozó kereslet mértéke és a kettő szorzatából a várható maximális árbevétel.

Az elemzés végén magától értetődő kérdés, hogy eredményeink mennyire megbízhatóak. A lineáris regresszióhoz hasonlóan a Tobit-modellnél is alkalmazhatóak olyan tesztek, amelyek az illeszkedés jóságát, a becslés torzítatlanságát és a homoszkedaszticitás feltételét vizsgálják. Ezeket a modellspecifikációnál nem részleteztem, de van egy nagyon egyszerű és egyértelmű mutatója a modellezés eredményességének: a találatmátrix. Ennek kiszámolásához egy új változóban rögzítettük a vásárlás becsült valószínűségét (y*), majd ebből képeztünk egy dummy változót, amelynek két értéke a nem vásárlók ($y^*<0,5$) és a vásárlók ($y^*\geq0,5$). A végeredmény egy kereszttábla lesz, amely összehasonlítja a kereslet becsült és a tényleges, az interjúalanyok által kinyilvánított értékeit. Kutatásunkban a helyes becslések aránya 78%, ami sajnos megerősíti azt az axiómát, hogy nincs tökéletes modell, és felveti az első- és másodfajú hiba költségkalkulációjának a szükségességét is.

Irodalomjegyzék

Breen, Richard 1996. *Regression Models*. Censored, Sample Selected, or Truncated Data. Sage University Paper

Greene, William 2003. *Econometric analysis.* Pearson Education International, Prentice Hall, Upper Saddle River, USA

Greene, William 1998. *Marginal Effects in the Censored Regression Model.* Economic Letters, Harvard University

Hausman, Jerry 2001. Microeconometrics. *Journal of Econometrics*, Volume 100, Issue 1, 2001/1, 33–35.

Long, J. Scott 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications

Ramathan, Ramu 2003. *Bevezetés az ökonometriába alkalmazásokkal.* Panem, Budapest