

Statisztika tanszék

TÉZISGYŰJTEMÉNY

Oravecz Beatrix

**SZELEKCIÓS TORZÍTÁS ÉS CSÖKKENTÉSE AZ
ADÓSMINŐSÍTÉSI MODELLEKNÉL**

című Ph.D. értekezéséhez

Témavezető:

Dr. Hunyadi László
kandidátus, egyetemi tanár

© Oravecz Beatrix

Tartalomjegyzék

| | |
|-------------------------------------------------------------|----|
| I. Kutatási előzmények és a téma indoklása | 4 |
| II. A felhasznált módszerek | 7 |
| III. Az értekezés eredményei | 9 |
| IV. Főbb hivatkozások..... | 11 |
| V. A témakörrel kapcsolatos saját publikációk jegyzéke..... | 18 |

I. Kutatási előzmények és a téma indoklása

Az elmúlt évtizedben tovább folytatódott a számítástechnikai kapacitások ugrásszerű növekedése. Ez az egymással versenyző, tömegtermelésre berendezkedett nagyvállalatok piacán egyrészt azt jelentette, hogy a cégek az üzleti folyamataikat leíró adatokat ma már elsősorban nem papíralapon, hanem elektronikus formában ún. adattárházakban tárolják. Másrészt lehetővé vált az így felhalmozódott adattömeg üzleti célú felhasználása elemzési, adatbányászati technikák segítségével a jobb üzleti döntések érdekében. Az adattömegben rejlő információk kinyerésére a software-fejlesztők különféle statisztikai/adatbányászati programcsomagokat fejlesztettek ki, melyek ma már széles körben elterjedtek.

Az adatok ilyen célú felhasználásában a pénzügyi szektor szereplői az élen járnak. Az élesedő verseny a szektor szereplőit rákényszerítette arra, hogy a rendelkezésükre álló erőforrásokat minél hatékonyabban használják fel. Jellemző példa erre a bankok esetében a lakossági (retail) hitelezési folyamatok automatizálása, melyek segítségével gyorsabban és lényegesen olcsóbban történhetnek a hitelezéssel kapcsolatos döntések. Az automatizált döntési rendszerek alkalmazásával lehetőség nyílt a hitelportfólió növelésére a hitelezési kockázat¹ kontrollja mellett. Ez a gyakorlatban azt jelenti, hogy a hitelező bankok minél több és jobb minőségű adatot próbálnak az ügyfelekről gyűjteni annak érdekében, hogy ezekből különböző adatbányászati módszerek segítségével minél több információhoz jussanak az ügyfelek fizetési képességével és hajlandóságával kapcsolatban. Ezt a célt szolgálja az automatizált adósmínősítési eszköz az ún. credit scoring, mely statisztikai módszerek segítségével képes megbecsülni az egyes adósok nemfizetési kockázatát. A módszer elterjedésének, az üzleti megfontolásokon túlmenően, a másik hajtóereje a Bázel II. tőke megfeleléssel kapcsolatos szabályozói direktiva, mely előírja az alkalmazó bankok számára, hogy az egyes adósok esetében a bank képes legyen az ügyfél várható nemfizetési

¹ A hitelezési kockázat annak a veszélyét fejezi ki, hogy a kölcsön adott tőkét és/vagy kamatait a hitelfelvevő részben vagy egészében nem fizeti vissza, és emiatt a bankot veszteség éri.

valószínűségét éves időhorizonton előrejelezni. A credit scoring modellek erre a problémára is megoldást kínálnak.

A credit scoring módszerek széleskörű alkalmazásának ellenére még mindig vannak a módszertannak olyan aspektusai, amelyek nem kaptak elegendő figyelmet sem a szakirodalomban, sem a gyakorlatban. A modellépítési minta reprezentativitásának kérdése például ilyen terület. Az adóminősítési modellek általában nem véletlen mintán épülnek, hiszen itt tipikusan csak azoknál az ügyfeleknél rendelkezünk teljes adatállománnyal, akik átestek egy hitelbírálati folyamaton és elfogadták őket. A kérelmek elfogadására/ elutasítására használt credit scoring modell idővel elveszti aktualitását, pontosságát, ezért újra kell építeni. Ha nem frissítik a modellt, akkor nem követi a populációban és a magyarázó változók hatásában bekövetkező változásokat, és az eredeti modell elveszíti prediktív erejét. Másrészt viszont, ha csak a befogadott ügyfelek adatait használják a modell frissítéséhez, akkor megkérdőjelezhető lesz az új modell érvényessége, hiszen a befogadottak és az elutasítottak eloszlása valószínűleg különbözik a szisztematikus elbírálási folyamat eredményeként, így a befogadottak nem reprezentálják a teljes sokaságot jelentő összes kérelmezőt. Ezt a jelenséget nevezzük elutasítási torzításnak (*reject bias*), vagy általánosabban *szelekciós torzítás*nak.

A dilemmára az elutasítottak jellemzőinek felhasználásával történő modellépítés (*reject inference*) jelenthet választ. Ez tulajdonképpen annak becslése, hogy hogyan viselkedett volna az elutasított kérelmező, ha megkapta volna a hitelt.

Egy gyakran idézett példa a büntetett előélet. A büntetett előéletű kérelmezőket majdnem mindig elutasítják. Ha mindet elutasítanak, akkor *reject inference* nélkül a végső modellben nem jelenne meg ez az ismérv. Az a tény, hogy a többséget elutasítják, gyakran azt jelenti, hogy a kisebbség, akit elfogadnak, nagyon speciális tulajdonságokkal rendelkezik, és egyáltalán nem reprezentálja a büntetett előéletűeket általában. Így, ha csak az elfogadottak teljesítését modellezzük, akkor a végső modellünk túlzottan optimista lesz.

A dolgozatban *a credit scoring modelleknél fellépő szelekciós torzítás csökkentésére* használható módszerekkel foglalkoztam. A jelenség vizsgálata a magyaryelvű szakirodalomból szinte teljesen hiányzik, csak említés szintjén találkozhattunk vele. A témaválasztást elméleti érdekességén túl *gyakorlati jelentősége* is indokolta. Hiszen ha csak egy kicsit is sikerül javítani a modellek teljesítményén, az óriási profitnövekedést, és/vagy kockázatcsökkenést eredményezhet *a bankok* számára, mivel nagy volumenű kihelyezésekről van szó. A kockázatok pontosabb értékelése ugyanakkor *az ügyfelek számára* is előnyös, mert a jó adósok számára a kockázati felár csökkentését teszi lehetővé, vagy megfelelő kockázati felárral olyanok is kaphatnak hitelt, akiket eddig elutasítottak.

A témához kapcsolódó korábbi kutatások alapján elmondható, hogy az elutasítottak alkalmazása a modellépítés során csak akkor lehet értelmes és hasznos megoldás, ha *bizonyos feltételek teljesülnek* az elfogadott és az elutasított sokaságra. A gyakorlatban működhetnek ezek a megoldások, mert a feltételezések sokszor indokoltak, vagy legalábbis jó irányba mutatnak. Például ésszerű feltételezés, hogy a rosszak aránya nagyobb az elutasítottakon belül, mint az elfogadottakon belül (azonos score mellett is), még akkor is, ha nem tudjuk korrekten számszerűsíteni, hogy mennyivel nagyobb.

Az elutasítottak tényleges és imputált adatainak alkalmazásának haszna függ az elutasítási aránytól, a mintabeli és sokasági eloszlásoktól és az alkalmazott statisztikai feltételek teljesülésétől. Van néhány portfólió, ahol nagyon alacsony az elutasítottak aránya (ilyen például a jelzáloghitelek piaca). Ilyen esetekben felesleges lehet az elutasítottakkal foglalkozni, hiszen elhanyagolható az arányuk a populáción belül, így az általuk okozott torzítás sem igényel korrekciót. Másrészt a nagyobb kockázatú portfóliók esetén, például a kis - és kezdő vállalkozások hitelezésénél, az elutasítási arány igen nagy lehet, így a szelekciós torzítást már nem lehet figyelmen kívül hagyni.

II. A felhasznált módszerek

A dolgozatban többféle módszer alkalmazhatóságát vizsgáltam, ezekből választottam aztán a konkrét kutatási adatbázis esetében használható, legjobbnak tűnő módszereket.

Az adósminősítési modelleknél fellépő szelekciós torzítás adathiányból eredő probléma, hiszen a korábban elutasított banki ügyfelek esetén a hitelkockázatot (hitelvisszafizetést) leíró változó értéke hiányzik (nem megfigyelhető), ezért először a *hiányzó adatok* típusait és kezelésük lehetséges módjait mutattam be.

A következő részben röviden áttekintettem a *credit scoring* feladatát és a leggyakrabban alkalmazott módszereket, valamint az ezek értékeléséhez használható mérőszámokat. A gyakorlatban az adósminősítési modellek esetében a **logisztikus regresszió** alkalmazása a legelterjedtebb, ezért az empirikus kutatás során én is logit modellt használtam az ügyfelek nemfizetési valószínűségének becslésére.

A III. részben ismertettem a szakirodalomban fellelhető *módszereket, amelyek a scoring modelleknél fellépő szelekciós torzítás csökkentését szolgálják*. Mindegyik módszer valamilyen módon felhasználja az elutasítottokról meglévő információkat.

Az elutasítottak tényleges visszafizetési adatát nem ismerjük, ezért – mivel a semmiből nem keletkezhet új információ-, ha fel akarjuk használni őket a modellépítéshez, akkor vagy *feltételezésekkel* kell élnünk, vagy *pótlólagosan információt* kell szerezni a visszafizetési viselkedésükről.

Bemutattam ezen (reject inference) technikák elméleti hátterét, kiemelve az alkalmazott feltételezéseket vagy a pótlólagos információ szerzésének és felhasználásának módját és összegeztem az eddigi gyakorlati tapasztalatokat.

Az alkalmazott feltételek teljesülése azonban általánosságban nem tesztelhető, így - a szakirodalom áttanulmányozása után - arra a következtetésre jutottam, hogy *a torzítás kiküszöbölésének egyetlen robusztus és megbízható módja, ha az*

elutasítottak egy részét ténylegesen meghitelezik és így figyelik meg viselkedésüket és esetleges bedőlésüket.

*Pótlólagos információk felhasználásával minden szempontból javítani tudunk a modellen, hiszen ekkor valóban több információra támaszkodunk a modellépítés során. Ezt az utat azonban nem mindig lehet megvalósítani, a megoldás pénz- és időigényes volta miatt. Az eljárás költségei csökkentésének egy lehetséges módja az általam **résnyire nyitott kapu**-nak nevezett módszer alkalmazása egyfajta költségoptimális mintaelosztással.*

Ez azt jelenti, hogy minden egyébként elutasítandó ügyfélnek van esélye a mintába kerülésre, de nem egyforma valószínűséggel. Kis valószínűséggel kaphatnak hitelt azok akiknél nagyobb a várható veszteség és nagyobb valószínűséggel azok akiknél ez a várható veszteség kisebb. Így egy rétegzett mintát kapunk egyfajta költségoptimális mintaelosztással. Végül **átsúlyozással** kaphatunk egy a sokaságot valóban reprezentáló mintát anélkül, hogy vállalni kellett volna a mindenki beengedésével járó hatalmas költségeket.

Az empirikus kutatás keretében egy valós banki adatbázison (lakossági hitelkártya adatokon) vizsgáltam egy logisztikus regresszióval épített scoring modellen a résnyire nyitott kapu módszerrel elérhető javulást, annak költségeit és várható hasznait.

III. Az értekezés eredményei

A dolgozattal ráirányítottam a figyelmet a szelekciós torzítás figyelembe vételének fontosságára, és a különböző torzítás csökkentő módszerek esetében a gyakorlati alkalmazás lehetőségeire.

A nemzetközi szakirodalom feldolgozásakor az adóminősítési modelleknél fellépő szelekciós torzítás csökkentésére alkalmazható módszereket az adathiány mechanizmusoknak megfelelően kategorizálva mutattam be.

Az *empirikus kutatás* keretében egy valós banki adatbázison (lakossági hitelkártya adatokon) vizsgáltam a résnyire nyitott kapu módszerrel elérhető javulást, annak költségeit és várható hasznait. Az elvégzett vizsgálat és modellszámítások eredményeként a gyakorlati szakemberek számára hasznosítható ajánlásokat fogalmaztam meg.

Fontos megállapítások:

- Az empirikus kutatás során azt tapasztaltuk, hogy *magas elutasítási arány (erőteljes és nem teljesen véletlenszerű szelekció) esetén gyengébb teljesítményű modellek építhetők*, mint kisebb arányú elutasítás esetén. Ennek egyik oka, hogy ekkor kevés rossz ügyfél kerül a portfólióba, ami megnehezíti a modellek számára a rosszak karakterisztikáinak megismerését. Másik oka, hogy a szelekció hatására egyébként szignifikáns magyarázó változók bizonyos értékei nem kerülnek a mintába, aminek következtében a magyarázó változó már nem lesz szignifikáns.
- Ilyen esetekben segíthet a pótlólagos információ szerzés egyik módja, ha belső forrásból, a résnyire nyitott kapu alkalmazásával nyerünk új megfigyeléseket. Azt láttuk, hogy a *nyitott kapu módszerrel javult a modellek teljesítménye, és ennek következtében a terméken elérhető profit is nőtt.*

- Azt tapasztaltuk, hogy ha a profitmaximalizálás a cél, akkor *jobb, ha az elméleti úton meghatározott cutoff értéket használjuk*, szemben a gyakorlatban elterjedt empirikus meghatározási móddal.
- Eredményeink szerint a modelljavulás és a profitnövekedés mértéke az első lépcsőben volt a legnagyobb. Tehát *leginkább az egyébként befogadandókhoz közel álló, azoktól csak kicsit rosszabbnak tűnő ügyfelekből érdemes résnyire nyitott kapuval beengedni még ügyfeleket*.

Ez az elsőlépcsős nagymértékű modelljavulás és profitnövekedés valószínű csak az adatbázis sajátossága, de egyéb, általános érvényű megfontolások is ezt a stratégiát sugallják. Az elfogadási tartományhoz közeli esetekre még jobbak a becsléseink. Ide még valószínűleg jól tudjuk becsülni a rosszak arányát, ezáltal a plusz minta költségei tervezhetőbbek, és kisebbek is, mintha egy távoli tartományból vennénk mintát.

Végezetül elmondható, hogy a dolgozatban ismertetett technikák és elméleti -, gyakorlati megfontolások nem csak a banki adósminősítés területén hasznosak és alkalmazhatók, hanem sok más olyan adatbányászati probléma esetén is, amelyek hasonló mintaszelekciós mechanizmust tartalmaznak.

IV. Főbb hivatkozások

E.I. ALTMAN [1968], "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy", *Journal of Finance*, 1968 szeptember, Vol.23. 589-609.old.

L. ASCH [1995], "How the RMA/Fair, Isaac Credit-Scoring Model Was Built," *Journal of Commercial Lending*, 10-16.o., 1995. június.

D. ASH ÉS S. MEESTER [2002], "Best Practices in Reject Interference," *Presentation at Credit Risk Modeling and Decision Conference*, Wharton Financial Institutions Center, Philadelphia, 2002. május.

T. ASTEBRO ÉS I. BERNHARDT [2001], "Bank Loans as Predictors of Small Start-up Business Survival", manuscript

T. ASTEBRO ÉS G. CHEN [2000], "Missing Data Analysis for Single Choice and Multiple Choice Survey Questions When Data are Sparse," *Presented at 2000 Academy of Management Conference*, Symposium entitled "Much ado about missing data," Kanada, Toronto.

S. AZEN ÉS M. VAN GUILDER [1981], "Conclusions regarding algorithms for handling incomplete data" *Proc. Stat. Computing Sec., Am. Statist. Assoc.*,53-56.o

J.B. BANASIK, J.N. CROOK ÉS L.C. THOMAS [2003], "Sample Selection Bias in Credit Scoring Models," *Journal of the Operational Research Society*, 54. szám 822-832.o.

J. BARNARD,ÉS D.B. RUBIN [1999], "Small-sample degrees of freedom with multiple imputation", *Biometrika*, 86. évf, 949-955.o.

H. BIERMAN ÉS W. H. HAUSMAN [1970], „The Credit Granting Decision,” *Management Sciences*, 16. évf., 519-532.o.

BÓDY SÁNDOR, SULYOK PAP MÁRTA [1997], "Cégminősítés", *Nemzetközi Bankárképző Központ*, Budapest

W.J. BOYES, D.L. HOFFMAN ÉS S.A. LOW [1989], "An Econometric Analysis of the Bank Credit Scoring Problem," *Journal of Econometrics*, 40. évf. 3-14.o.

G. CHEN ÉS T. ASTEBRO [2003], "How to Deal with Missing Categorical Data: Test of a Simple Bayesian Method," *Organizational Research Methods*, 6. évf. 3. szám 309-321.o.

G. CHEN ÉS T. ASTEBRO [2003], "Bound and Collapse Bayesian Reject Inference When Data are Missing not at Random," in T. Astebro, P. Beling, D. Hand, B. Oliver és L.B. Thomas (Eds.): *Mathematical Approaches to Credit Risk Management, Conference Proceedings, Banff International Research Station for Mathematical Innovation and Discovery*, 2003. október 11-16.

G. CHEN ÉS T. ASTEBRO [2006], "A Maximum Likelihood Approach for Reject Inference in Credit Scoring", November 25, 2006, manuscript, available at SSRN: <http://ssrn.com/abstract=872541>

R.K. CHHIKARA [1989], "The State of the Art in Credit Evaluation", *American Journal of Agricultural Economics*, 71.évf, 5.szám, 1138-1144.o.

J.B. COPAS ÉS H.G. LI [1997], "Inference for Non-random Samples (with discussion)," *Journal of the Royal Statistical Society*, B, 59. évf. 55-95.o.

J. CROOK és J. BANASIK [2002], "Does Reject Inference Really Improve the Performance of Application Scoring Models?" *Working paper 02/3*, Credit Research Centre, University of Edinburg, Anglia.

A.P. DEMPSTER, N.M. LAIRD és D.B. RUBIN [1977], "Maximum Likelihood Estimation From Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society B*, 39. évf. 1-38.o.

S.G. DONALD [1995], "Two-step Estimation of Heteroskedastic Sample Selection Models," *Journal of Econometrics*, 65. évf. 347-380.o.

A.J. FEELDERS, S. CHANG és G.J. MCLACHLAN [1998], "Mining in the Presence of Selectivity Bias and its Application to Reject Inference," *Proceedings of the fourth international conference on knowledge discovery and data mining (KDD-98)*, AAAI Press, 199-203.o.

A.J. FEELDERS [1999], "Credit Scoring and Reject Inference with Mixture Models," *International Journal of Intelligent System in Accounting, Finance and Management*, 8 évf. 271-279.o.

A.J. FEELDERS [2000], "Credit Scoring and Reject Inference with Mixture Models," *International Journal of Intelligent System in Accounting, Finance and Management*, 9. évf. 1-8.o.

A.J. FEELDERS [2001], "An Overview of Model Based Reject Inference for Credit Scoring," *Working paper*, Institute for Information and Computing Sciences, Utrecht University, Hollandia.

D.J. FORGARTY [2005], "Multiple Imputation as a Missing Data Approach to Reject Inference on Consumer Credit Scoring", *Manuscript*, DavidF1967@email.uopx.edu

H. FRIEDMAN, E. ALTMAN ÉS D.L. KAO [1985], "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress", *Journal of Finance*, 1985. március, Vol. 40. Iss.1, 269-291 old.

A. GELMAN, J.B. CARLIN, H.S. STERN ÉS D.B. RUBIN [1995], "Bayesian Data Analysis", *Chapman & Hall*.

W.R. GILKS, S. RICHARDSON, ÉS D. J. SPIEGELHALTER (Eds.) [1996]. „, Markov Chain Monte Carlo in Practice”, *Chapman & Hall*, London.

R. GLYNN, N. M. LAIRD ÉS D.B. RUBIN [1986], „Selection modeling versus mixture modeling with nonignorable nonresponse”, *In H. Wainer (ed.) Drawing Inferences from Self-Selected Samples*, 119-146. New York: Springer-Verlag.

J.W. GRAHAM ÉS S.I. DONALDSON [1993], "Evaluating Interventions with Differential Attrition: the Importance of Nonresponse Mechanisms and Use of Followup Data," *Journal of Applied Psychology*, 78. évf. 119-128.o.

W.H. GREENE [1998], "Sample Selection in Credit-scoring Models," *Japan and the World Economy*, 10. évf. 299-316.o.

HADU OTTÓ [2003]: „Többváltozós statisztikai számítások”, KSH, Budapest

HAJDU OTTÓ ÉS VIRÁG MIKLÓS [1996], „Pénzügyi mutatószámokon alapuló csődmodell-számítások”, *Bankszemle*, 1996/1-2

HÁMORI GÁBOR [2001], „A CHAID alapú döntési fák jellemzői”, *Statisztikai szemle*, 79évf. 8.szám, 703-710.o.

D.J. HAND és W.E. HENLEY [1993/4], „Can Reject Inference Ever Work?,” *IMA Journal of Mathematics Applied in Business & Industry*, 5. évf. 4. szám 45-55.o.

D.J. HAND és W.E. HENLEY [1994], „Inference About Rejected Cases in Discriminant Analysis,” *Springer*, 292-299.o., New York.

D.J. HAND [1997], „Construction and Assessment of Classification Rules,” *Chichester: Wiley*.

D.J. HAND [1998], „Reject Inference in Credit Operations,” in *Credit Risk Modeling: Design and Application* (ed. E. Mays), 181-190.o. AMACOM.

D.J. HAND [2001], „Measuring Diagnostic Accuracy of Statistical Prediction Rules,” *Statistica Neerlandica*, 53. évf. 3-16.o.

J.J. HECKMAN [1979], „Sample Selection Bias as a Specification Error,” *Econometrica*, 47. évf. 153-161.o.

D. HEDEKER és R. D. GIBBONS [1997], „Application of random-effects pattern-mixture models for missing data in longitudinal studies”, *Psychological Methods*, 2(1), 64-78.

C.W. HOLSAPPLE et.al. [1988], „Adapting Expert System Technology to Financial Management” *Financial Management*, 1988. ősz, 19.évf. 12-22.o.

D.C HSIA [1978], „Credit Scoring and the Equal Credit Opportunity Act,” *The Hastings Law Journal*, 30. évf. 371-448.o., 1978. november.

HUNYADI LÁSZLÓ [2001], „A mintavétel alapjai”, *Egyetemi Jegyzet SZÁMALK*, Budapest

HUNYADI LÁSZLÓ és VITA LÁSZLÓ [2002], „Statisztika közgazdászoknak”, *Központi Statisztikai Hivatal*, Budapest

T. JACOBSON és K.F. ROSZBACH [1999], „Evaluating Bank Lending Policy and Consumer Credit Risk,” in *Computational Finance 1999* [edited by Y.S. Abu-Mostafa et al.] the MIT Press, 2000.

D.N JOANES [1993], „Reject Inference Applied to Logistic Regression for Credit Scoring,” *IMA Journal of Mathematics Applied in Business and Industry*, 5. évf. 4.szám 35-43.o.

N.M. KIEFER és C.E. LARSON [2003], "Specification and Informational Issues in Credit Scoring", kézirát

J.O. KIM és J. CURRY [1977], „The treatment of missing data in multivariate analysis”, *Sociol. Meth. Res.*, 6.évf, 215-240.o.

P.W. LAVORI, R. DAWSON és D. SHERA [1995], „A multiple imputation strategy for clinical trials with truncation of patient data”, *Statistics in Medicine*, 14 évf., 1913-1925.o.

R.J.A. LITTLE [1979], "Maximum Likelihood Inference for Multiple Regression with Missing Values: A Simulation Study", *Journal of the Royal Statistical Society*, vol.41., 76-87.o.

R.J.A. LITTLE és D.B. RUBIN [1987], "Statistical Analysis with Missing Data", *John Wiley & Sons*, New York.

R.J.A. LITTLE és D.B. RUBIN [2002], "Statistical Analysis with Missing Data," 2. Edition, *John Wiley & Sons*, New York.

R.J.A. LITTLE [1993], "Pattern-mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association*, 88. évf. 125-134.o.

R.J.A. LITTLE és N. SCHENKER [1994], "Missing Data" in *Handbook for Statistical Modeling in the Social and Behavioral Sciences* [G. Arminger, C. C. Clogg és M. E. Sobel szerk.] New York: Plenum 39-75.o.

G.S. MADDALA [1983], "Limited Dependent and Qualitative Variables in Econometrics," *Cambridge University Press*, Cambridge, UK.

MÁDER MIKLÓS PÉTER [2005], "Imputálási eljárások hatékonysága", *Statisztikai Szemle*, 83.évf. 7.szám, 628-644.o.

N.K. MALHOTRA [1987], "Analyzing Marketing Research Data with Incomplete Information on the Dependent Variable", *Journal of Marketing Research*, vol.24, 74-84.o

O.L. MANGASARIAN [1965], "Linear and nonlinear separation of patterns by linear programming", *Operation Research*, 13 évf., 444-452.o.

E. MAYS [2004], "Credit Scoring for Risk Managers" *South Western Thomson Learning*

G.J. MCLACHLAN és K.E. BASFORD [1988], "Mixture Models, Inference and Applications to Clustering," *Marker Dekker*, New York.

G.J. MCLACHLAN [1992], "Discriminant Analysis and Statistical Pattern Recognition," *Wiley*, New York.

R.W. MCLEOD et.al. [1993], "Predicting Credit Risk: A Neural Network Approach", *Journal of Retail Banking*, 15.évf., 3.szám, 37-40.o.

C.L. MENG és P.SCHMIDT [1985], "On the Cost of Partial Observation in the Bivariate Probit Model," *International Economic Review*, 26. évf. 1. szám 71-85.o., 1985. február.

M.C. PAIK, R. SACCO és I.F. LIN [2000], "Bivariate Binary Data Analysis with Nonignorably Missing Outcomes," *Biometrics*, 56. évf. 1145-1156.o.

D.J. POIRIER [1980], "Partial Observability in Bivariate Probit Model," *Journal of Econometrics*, 12. évf. 209-217.o.

P.L. ROTH és F.S. SWITZER III [1995], "A Monte Carlo Analysis of Missing Data Techniques in a HRM Setting," *Journal of Management*, 21. évf. 5. szám 1003-1023.o.

D.B. RUBIN [1987], "Multiple Imputation for Nonresponse in Surveys," *John Wiley & Sons*.

D. B. RUBIN [1996], "Multiple imputation after 18+ years [with discussion]," *Journal of the American Statistical Association*, 91, 473-489.o.

D. B. RUBIN [2003], "Nested Multiple Imputation of NMES via Partially Incompatible MCMC", *Statistica Neerlandica*, 57, 3-18.o.

RUDAS T. [1998], "Hogyan olvassunk közvélemény-kutatásokat?", Új Mandátum Könyvkiadó, Budapest

J.L. SCHAFER és J.W. GRAHAM [2002], "Missing Data: our View of the State of the Art," *Psychological Methods*, 7. évf. 2. szám 147-177.o.

- J.L. SCHAFER és M.K. OLSEN [1998], "Multiple imputation for multivariate missing-data problems: a data analyst's perspective", *Multivariate Behavioral Research*, 33. évf., 545-571. o.
- P. SEBASTIANI és M. RAMONI [2000], "Bayesian Inference with Missing Data Using Bound and Collapse," *Journal of Computational and Graphical Statistics*, 9. évf. 4. szám 779-800.o.
- K.Y. TAM [1991], "Neural Network Models and the Prediction of Bank Bankruptcy" *Omega. The International Journal of Management Science*, 19.évf., 5.szám 429-445.o.
- K.Y. TAM és M.Y. KIANG [1992], "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions". *Management Science*, 1992. július, 38.évf. 7.szám, 926-947o.
- L.C. THOMAS [2000], "A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers," *International Journal of Forecasting*, 16. évf. 149-172.o.
- L.C. THOMAS, D.B. EDELMAN és J.N. CROOK [2002], "Credit Scoring and Its Applications", *Society for Industrial and Applied Mathematics, Philadelphia*
- F. VELLA [1992], "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models", *Journal of Applied Econometrics*, 7. évf. 413-421.o.
- VIRÁG MIKLÓS és KRISTÓF TAMÁS [2006], "Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel", *Vezetéstudomány*, 37. évf., 1.szám. 25-35.o.
- D. WEST [2000], "Neural Network Credit Scoring Models", *Computers and Operations Research*, 27-évf. 1131-1152.o
- J.C. WIGINTON [1980], "A note on the comparison of logit and discriminant models of consumer credit behaviour", *Journal of Financial Quantitative Anal.*, 15, 757-770.o.

V. A témakörrel kapcsolatos saját publikációk jegyzéke

Folyóiratcikkek:

Oravecz Beatrix (2007): Credit scoring modellek és teljesítményük mérése. Hitelintézeti Szemle, VI.évf., 6.szám, 607- 627.o

Oravecz Beatrix : Hiányzó adatok és kezelésük a statisztikai elemzésekben. Statisztikai Szemle, várható megjelenés 2008. január

Oravecz Beatrix: Szelekciós torzítás és csökkentése a credit scoring modelleknél. Hitelintézeti Szemle, várható megjelenés 2008. január

Cikk szerkesztett könyvben:

Oravecz Beatrix (2004): Imputációs eljárások. Egy reneszánsz statisztikus. Tanulmánykötet Hunyadi László tiszteletére, KSH

Egyéb:

Oravecz Beatrix (2000): Hitelmonitoring, Szakdolgozat, BKAE Pénzügy Tanszék

Oravecz Beatrix (2002): Dijkstra, W.: Új módszer az interjúk közbeni kölcsönkapcsolatok tanulmányozására. cikkismertetés, Statisztikai szemle, 79. évf. 7. szám, 636-637.o.