

STOCK ÉS FLOW TÍPUSÚ SZÁMVITELI ADATOK ALKALMAZÁSA A CSŐDELŐREJELZŐ MODELLEKBEN

A vállalatok hitelkockázatának megítélését szolgáló scoring modellekben általánosan elterjedt a számviteli adatok alapján kalkulált hányados típusú pénzügyi mutatószámok alkalmazása. E változók körében azonban vannak olyanok, amelyek mérlegtételeket (mint stock típusú adatokat) és eredménykimutatásból vett értékeket (mint flow típusú adatokat) vetnek össze. Az ilyen mutatók esetén a stock típusú adatokat a nyitó és záró érték átlagában szükséges figyelembe venni. Ennek ellenére a tudományos kutatásban és a gyakorlati modellezésben is gyakori, hogy az elemzők a stock típusú adatok vonatkozásában elmulasztják az átlagolást. A tanulmány azt a kérdést vizsgálja, hogy e „mulasztásnak” van-e statisztikailag mérhető hatása a modellek előrejelző képességére. Az empirikus vizsgálat során a logisztikus regresszió módszerét alkalmaztam tízszeres keresztvalidációval. A kutatás eredménye arra utal, hogy az időszak végi értékek használata az átlagok helyett csak kis – a vizsgált adathalmaz esetén egy százalékpontonál kisebb – mértékben csökkenti a modellek előrejelző képességét a ROC-görbe alatti területet használva teljesítménymutatóként.

Kulcsszavak: csődelőrejelzés, pénzügyi mutatók, stock és flow adatok, logisztikus regresszió

A tanulmány a modern pénzügyek egyik vezető témájával, a hitelkockázati előrejelzéssel (Fantazzini – Figini, 2009) foglalkozik. Amendola et al. (2011) véleménye szerint e témakör ugyanolyan releváns és kutatott terület ma, mint az 1930-as években. Ennek legfőbb oka, hogy a hiteladósok nemfizetése jelentős veszteséget okozhat az üzleti élet szereplőinek, valamint a nemzetgazdaság egészének is (Horta – Camanho, 2013).

Fontos azonban hangsúlyozni, hogy ez a témakör nemcsak a hitelezési veszteségek minimalizálása szempontjából jelentős. Nemzetgazdasági relevanciája abban is rejlik, hogy a modellek előrejelző képességének növelésével makroszinten is fokozható a forrásallokáció hatékonysága (Min – Jeong, 2009). Mikroszinten az az aspektus említhető, hogy a modellek alkalmazásával részben automatizálható a hitelezési döntéshozatal, ugyanis a modellek segítségével részletes elemzés nélkül azonosítható a hitelezésre nem ajánlott ügyfélkör, így a döntéselőkészítés munkaiigénye csökken (Ic – Yurdakul, 2010). A bázeli szabályozásból adódóan a hitelkockázati modellek fejlesztése mikro- és makroszinten is fontos tényezővé vált, mivel a tőkekövetelmény összegét jelentősen befolyásolja a bank belső minősítési rendszerében alkalmazott modell által becsült bedőlési valószínűség (Sánchez-Lasheras et al., 2012). Így a minél fejlettebb kockázatbecslési módszertan egyszerre szolgálja mikroszinten a bankok egyéni profitmaximalizálási célját és nemzetgazdasági szinten a bankrendszer stabilitását.

A hitelkockázati modellezésben meg kell különböztetni a lakossági, a vállalati és az önkormányzati ügyfeleket abból a szempontból, hogy milyen információk állnak rendelkezésre, illetve bírnak magyarázó erővel az adott szegmens vonatkozásában. Ez a tanulmány a vállalatok jövőbeli fizetőképességével kapcsolatos kockázat megítélésére alkalmazható modellek építésével foglalkozik. Az e körbe tartozó modelleket a nemzetközi szakirodalom gyakran csődelőrejelző modelleknek nevezi. A csődelőrejelzés tu-

dományterületének alapfeltevése Zhou (2013) szerint az, hogy a csőd (vagy default) előtti év pénzügyi adatai és a csődesemény bekövetkezése közt kapcsolat van. A csődelőrejelző modellek célja, hogy e kapcsolatot számszerűsítsék. A modellek gyakorlati alkalmazása során a feltárt összefüggést használják fel a modellépítési mintában nem szerepelt vállalatok jövőbeli fizetőképességének előrejelzésére.

Az elemző által relevánsnak vélt magyarázó változók és a fizetőképesség jövőbeli bekövetkezése közti kapcsolatot a gyakorlati modellezés során jellemzően statisztikai modellek segítségével tárják fel. A potenciális magyarázó változók köre széles, ugyanis egy vállalkozás jövőbeli csődjének bekövetkezését számos tényező befolyásolhatja. A lehetséges magyarázó változók közt a számviteli kimutatásokból kinyerhető információk domináns szerepet töltenek be a szakterület tudományos kutatásában és gyakorlatában egyaránt. Alkalmazásuk már a 60-as években általánossá vált és népszerűségük napjainkig töretlen (Chen, 2012). Ennek egyik legfőbb oka az egyszerűség (Emel et al., 2003) és a napjainkban a fokozódó digitalizációnak köszönhetően az egyre könnyebb hozzáférhetőség.

A számviteli adatokból számítható hányados típusú pénzügyi mutatók mellett másodlagos szerepet töltenek be az egyéb magyarázó változók. Fontos azonban kiemelni, hogy ezek sok esetben legalább annyira fontosak a modellek előrejelző képessége szempontjából. Ennek okait a számviteli adatok hátrányai közt találhatjuk meg. Ezek közül talán az egyik legfontosabb, hogy a számviteli kimutatásokban található információk elsősorban a múltban meghozott vezetői döntések hatásait tükrözik, így a jövőre vonatkozó előrejelző képességük kétségtelenül korlátozott (Lin et al., 2014). A számviteli információk további hátránya, hogy nem minden esetben nyújtanak teljes képet a vállalatok pénzügyi helyzetéről (például mérlegen kívüli tételek), ennek oka lehet az is, ha a vállalat vezetése ma-

nipulálja a beszámolóban közölt adatokat a számviteli törvény által lehetővé tett keretek közt, vagy akár azon túl. A hitelkockázati modellezés vonatkozásában további problémát jelenthet, hogy a kisebb vállalkozások esetén kérdéses lehet a beszámolóban szereplő adatok megbízhatósága abból adódóan is, hogy azokat nem szükséges könyvvizsgáltnak alávetni. E hiányosságokat azonban képesek lehetnek kiküszöbölni az úgynevezett nem pénzügyi mutatók (a vállalatvezetés végzettsége, tapasztalata, a vállalat korábbi hiteltörténete stb.), amelyek sok esetben fontos kvalitatív jellemzőket ragadnak meg, releváns kiegészítő szerepet betöltve a modellekben (Kim – Han, 2003).

A kvalitatív változók esetében azonban jelentős korlátot jelent az adatok hozzáférhetősége, melynek köszönhetően a számviteli adatok – az imént kiemelt hiányosságok ellenére is – domináns szerepet töltenek be a hitelkockázati modellezésben. A témakör szempontjából a pénzügyi-számviteli adatok relevanciája vitathatatlan, azonban a tudományos kutatásban gyakori, hogy e változókat a modellekben egyszerűen csak „betöltik”, mint szükséges input tényezőket és a klasszifikációs módszerek alkalmazásától, illetve azok matematikai hátterének korszerűsítésétől várják a jobb előrejelző teljesítményt. Egyetértve azonban Sun (2007) azon megállapításával, miszerint a csödelőrejelző modellek előrejelző képessége nem csak újabb klasszifikációs módszerek alkalmazásával javítható, hanem új változók képzésével is, ezért ebben a tanulmányban a hitelkockázati modellezésben általános elterjedt és elfogadott logisztikus regresszió keretei közt maradok. A cél nem valamely új klasszifikációs eljárás bemutatása, illetve más eljárásokkal való összevetése lesz, hanem annak vizsgálata, hogy egyes pénzügyi mutatószámok számításmódjának megváltoztatása befolyásolja-e a csödelőrejelző modellek előrejelző képességét. Tömören összefoglalva: azt a kérdést vizsgálom, hogy milyen hatást gyakorol a modellek teljesítményére és felépítésére, ha a stock és flow jellegű adatok összevetése során az állományi értékeket nem a nyitó és záró adatok átlagaként, hanem – egyszerűen csak – a tárgyév záró értékeként vesszük figyelembe.

A következő szakaszban bemutatom, hogy melyek azok a csödelőrejelzésben gyakran alkalmazott pénzügyi mutatók, amelyek stock és flow értékeket vetnek össze. A harmadik szakaszban az empirikus vizsgálathoz felhasznált adatbázist mutatom be, majd ezt követően a számítások eredményeit ismertetem. A tanulmányt egy összefoglaló szakasz zárja, ahol az empirikus vizsgálatok eredményeit és az azokból levonható következtetéseket összegzem, illetve kitérek a bemutatott kutatás korlátaira és az azokból fakadó lehetséges jövőbeli vizsgálati területekre.

Stock és flow adatok a pénzügyi mutatószámokban

A stock, azaz állományi adatok egy időpontra vonatkozóan értelmezettek, míg a folyam jellegű (flow) értékek egy időszak során felhalmozódott mennyiségeket jelentenek (Hunyadi – Vita, 2004). Mivel a mérleg egy vál-

lalkozás vagy helyzetét a mérleg fordulónapjára vonatkozóan tükrözi, az eredménykimutatásban található adatok pedig egy naptári évre vonatkoznak (Virág et al., 2013), így a mérlegtételek stock jellegű adatok, az eredménykimutatás értékei pedig flow jellegű információknak tekinthetők.

A vállalatok pénzügyi teljesítményének megítélésében leggyakrabban használt pénzügyi mutatókat Virág et al. (2013) részletesen tárgyalja. Az idézett szerzők kitérnek arra is, hogy ha egy mutató számlálóját és nevezőjét is azonos kimutatásból vesszük, akkor az adatok stock vagy flow jellegének nincs jelentősége. Abban az esetben azonban, ha a hányados két tagja különböző kimutatásból származik – azaz egy mérlegtételt és az eredmény kimutatás egy adott sorát vetjük össze – akkor az állományi adatokat a nyitó és záró érték átlagában szükséges figyelembe venni annak érdekében, hogy statisztikailag összevethető legyen egy flow jellegű adattal. Annak ellenére, hogy a számításmód az egyetemi tankönyvek anyagát képezi, a tudományos kutatás és a gyakorlati modellezés esetén gyakran tapasztalható, hogy az elemzők a stock értékek esetén az előbb említett átlagolást elmulasztva az eredménykimutatás értékeit az adott év mérlegéből vett záró értékekkel vetik össze.

A csödelőrejelzés hazai szakirodalma megosztott a kérdés szempontjából, ugyanis számos olyan kutatás található, ahol a szerzők a publikációjukban vagy nem térnek ki arra, hogy pontosan hogyan számították a pénzügyi mutatókat, vagy egyértelműen jelzik, hogy stock és flow adatok összevetésekor az állományi értékeket a vizsgált év záró értékén vették figyelembe. Néhány önkényesen kiragadott példa a teljesség igénye nélkül: Virág – Hajdu (2001), Virág – Kristóf (2005), Virág – Kristóf (2006), Bozsik (2010), Virág – Nyitrai (2013), Virág – Nyitrai (2014). Találhatók azonban olyan tanulmányok is, amelyek egyértelműen rögzítik, hogy a kérdéses mutatók esetén a stock adatokat átlagolva vették figyelembe. Szintén néhány önkéntesen kiragadott példa a teljesség igénye nélkül: Virág – Kristóf (2009), Kristóf – Virág (2012).

A nemzetközi szakirodalomban nincs tudomásom olyan tanulmányról, amelyben a szerzők a pénzügyi mutatók számítása során kitértek volna arra, hogy az állományi és folyam jellegű értékek összevetésekor a stock értékeket átlagolták-e. Ez egyrészt magyarázható azzal, hogy a szerzők evidenciaként kezelik az átlagolás szükségességét, így azt fel sem tűntetik a publikációkban, másrészt azonban előfordulhat az is, hogy nem végzik el az átlagolást, hanem az állományi adatokat a tárgyév záró értékén veszik figyelembe akkor, amikor flow értékekkel végeznek összevetést. Említést érdemel az is, hogy még egyetlen olyan tanulmánnyal sem találkoztam, ahol kitértek volna ennek szükségességére, vagy vizsgálták volna ezen átlagolás elmulasztásának hatását a csödelőrejelző modellek előrejelző képességére. Ez a felismerés motiválta az itt bemutatott empirikus vizsgálat elvégzését annak érdekében, hogy referenciát nyújtson a jövőben a tudományos kutatás és a gyakorlati hitelkockázati modellezés számára.

Az empirikus vizsgálatához felhasznált adatok

Ha a témakör gyakorlati relevanciáját vesszük alapul, akkor az empirikus vizsgálatban valós hitelkockázati adatbázist lenne célszerű felhasználni. Ilyen adathalmazhoz azonban nehéz hozzájutni (Santos Silva – Murteira, 2009), vagy ha még hozzáférhető is, az adatokat, illetve az azok alapján végzett kvantitatív vizsgálatok eredményeit üzleti okok miatt nem lehet publikálni. A kutatás során magam is ezzel a problémával szembesültem, így empirikus vizsgálataimat egy saját adatgyűjtés eredményeképp összegyűjtött minta adatain végeztem el.

Az elemzést Magyarországon működő társas vállalkozások adatain végeztem. A mintába azon cégek kerültek be, amelyek hirdetményt tettek közzé a Cégbőlönny véletlenszerűen kiválasztott számaiban. A fizetéseketelenség tényét jogi értelemben definiáltam: az a vállalkozás minősült csődösnek, amellyel szemben felszámolási eljárást vagy csődeljárást kezdeményeztek. Ennek tényét a Cégjegyzék hatályos adatai alapján ellenőriztem. Ebből adódóan működőnek azon cégek minősültek, amelyekkel szemben az előbb említett két eljárás közül egyikre sem került sor a mintavétel időpontjában. A mintába került vállalkozások beszámolóit az elektronikus beszámoló portálról¹ töltöttem le, ahol a hatályos jogszabályok értelmében minden hazánkban működő vállalkozás köteles elérhetővé tenni számviteli kimutatásait. Annak érdekében, hogy a felállítandó modellek valós előrejelző képességgel bírjanak, a csődös vállalkozások esetén a beszámolók adatait a felszámolási vagy csődeljárás megindítását közvetlenül megelőző, de még teljes üzleti évre vonatkozóan vettem figyelembe.

Tekintettel arra, hogy az előbb említett nyilvános adatbázisokban csak a vállalkozás neve, illetve cégjegyzékszám alapján van lehetőség keresésre, az adatgyűjtés során az alábbi szempontokat vettem figyelembe:

- A vállalatnak legalább négy egymást követő üzleti évre visszamenőleg legyenek elérhetőek a beszámolói. A szempont alkalmazását az indokolta, hogy kiszűrhetőek legyenek az induló vállalkozások, amelyek a kezdeti pénzügyi nehézségeik miatt gyakran jobban hasonlítanak a csődös vállalatokra, mint a működőkre (Du Jardin, 2010), így a működő vállalkozások közt történő szerepeltetésüknek torzító hatása lehetett volna a modellekre.
- A vállalatnak folyamatosan legalább négy évre visszamenőleg pozitív árbevételt kell realizálnia annak érdekében, hogy valós gazdálkodási tevékenységet végző vállalkozások kerüljenek a mintába. E szempont alól egyetlen kivételt azon csődös vállalatok jelentettek, amelyek csak a csődöt közvetlenül megelőző évben nem realizáltak árbevételt. A kivétel oka, hogy az adatgyűjtés során gyakran tapasztaltam, hogy a csődöt megelőző évben a magyar vállalkozások sok esetben nem realizálnak árbevételt. E kivétel nélkül véleményem szerint fontos csődelőrejelző tényező esett volna ki a modelltől.
- A vállalat számviteli adatai és pénzügyi mutatói mutassanak időbeli szóródást legalább a mintavétel

megelőző négyéves időszakban. Ritkán ugyan, de előfordultak olyan vállalkozások, amelyek beszámolóiban éveken át állandó értékek szerepelnek. Ilyen esetben azonban a tanulmányban vizsgált kérdésnek nincs relevanciája, ugyanis éveken át azonos stock értékek esetén nincs jelentősége annak, hogy átlagoljuk-e azokat a flow értékekkel való összevetés során.

1. táblázat

Az empirikus vizsgálatban felhasznált pénzügyi mutatók és azok számításmódja

Mutatószám neve	Számításmód
Likviditási ráta (X_1)	Forgóeszközök/Rövid lejáratú kötelezettségek
Likviditási gyorsráta (X_2)	(Forgóeszközök – Készletek)/Rövid lejáratú kötelezettségek
Pénzeszközök aránya (X_3)	(Pénzeszközök + Értékpapírok)/Forgóeszközök
Cash flow/Kötelezettségek (X_4)	(Adózás utáni eredmény + Értécsökkenési leírás)/Kötelezettségek átlagos állománya
Cash flow/Rövid lejáratú kötelezettségek (X_5)	(Adózás utáni eredmény + Értécsökkenési leírás)/Rövid lejáratú kötelezettségek átlagos állománya
Tőkeellátottság (X_6)	(Befektetett eszközök + Készletek)/Saját tőke
Eszközök forgási sebessége (X_7)	Értékesítés nettó árbevétele/Átlagos mérlegfőösszeg
Készletek forgási sebessége (X_8)	Értékesítés nettó árbevétele/Átlagos készletállomány
Követelések forgási ideje (X_9)	Követelések átlagos állománya/Értékesítés nettó árbevétele
Eladósodottság (X_{10})	Kötelezettségek/Mérlegfőösszeg
Saját tőke aránya (X_{11})	Saját tőke/Mérlegfőösszeg
Bonitás (X_{12})	Kötelezettségek/Saját tőke
Árbevétel-arányos nyereség (X_{13})	Adózás utáni eredmény/Értékesítés nettó árbevétele
Eszközarányos nyereség (X_{14})	Adózás utáni eredmény/Átlagos mérlegfőösszeg
Követelések/Rövid lejáratú kötelezettségek (X_{15})	Követelések/Rövid lejáratú kötelezettségek
Nettó forgótőke aránya (X_{16})	(Forgóeszközök – Rövid lejáratú kötelezettségek)/Mérlegfőösszeg
Vállalat mérete (1) (X_{17})	Az eszközállomány természetes alapú logaritmus
Vállalat mérete (2) (X_{18})	Az árbevétel természetes alapú logaritmus
Eredménytartalék aránya (X_{19})	Eredménytartalék/Mérlegfőösszeg

Az adatgyűjtés eredményeképp létrejött adatbázis 2094 vállalkozás adatait tartalmazza a 2009-2015-ös időszakra vonatkozóan. A vállalatok körében a csődbe mentek hányada 46, a működőké pedig 54%-os arányt tett ki. A közel egyenlő arányú mintafelosztás használata általánosan mondható a csődelőrejelzésben (Amendola et al.,

2011). Alkalmazásának leggyakoribb oka a minél nagyobb mintaméret iránti igény, ugyanis a reprezentativitás követelményének érvényesítése még egy ilyen relatíve nagyméretű adathalmaz esetén is azt eredményezné, hogy a csődös megfigyelések száma viszonylag alacsony lenne. Mivel a csődölőjelző modellek fejlesztésének elsődleges célja a csődös vállalatok minél hatékonyabb azonosítása (Du Jardin, 2010), így e cél eléréséhez magam is célszerűnek tartottam a közel azonos arányú mintafelosztás alkalmazását annak érdekében, hogy a modellépítés során a lehető legtöbb információ álljon rendelkezésre a csődös vállalkozásokra vonatkozóan.

A modellek lehetséges magyarázó változóit részben a csődölőjelzés hazai szakirodalmára építve, részben pedig saját megfontolások alapján válogattam. Az empirikus vizsgálat független változóit és azok számítás módját az 1. táblázat mutatja.

A táblázatban vastag betűvel emeltem ki a tanulmány szempontjából releváns változókat. Azok ugyanis stock és flow típusú adatokat vetnek össze, így a számítás során – statisztikai szempontból – a stock értékek átlagát szükséges venni. A tanulmányban a vastaggal kiemelt változók kétféle megközelítésben szerepelnek majd: i) egyrészt az 1. táblázatban bemutatott módon, ii) másrészt oly módon, hogy a mérlegtételek az adott évi záró érték formájában szerepelnek a mutatók számítása során. A kutatás célja annak vizsgálata, hogy a vastag betűvel kiemelt mutatók esetén a számítás mód megváltoztatása milyen hatást gyakorol a logisztikus regresszióval felállított csődmodellek előrejelző képességére.

A tanulmányban megfogalmazott kutatási kérdésre akár mélyebb tudományos vizsgálatok nélkül is választ kaphatunk, ugyanis abban az esetben, ha az érintett mutatószámok számításánál felhasznált nyitó és záró mérlegtételek közt nincs, vagy csak nagyon alacsony a különbség, akkor a mutatók értéke kapcsán nincs relevanciája a stock és flow típusú adatok megfelelő kezelésének. Ennek érdekében az empirikus vizsgálatok elvégzését megelőzően kiszámítottam az érintett mérlegtételek vonatkozásában a nyitó és záró értékek változásának átlagát a működő és a csődös vállalatok vonatkozásában külön-külön. Ezek eredményeit szemlélteti a 2. táblázat.

2. táblázat

Az empirikus vizsgálatban releváns mérlegtételek változásának átlaga záró/nyitó viszonylatban

Vállalat	Kötelezettségek	Rövid lejáratú kötelezettségek	Mérlegfőösszeg	Készletek	Követelések
Működő	+23,6 %	+33,2 %	+9,1 %	+7,5 %	+27,0 %
Csődös	+6,3 %	+16,8 %	-32,0 %	-29,2 %	-9,2 %

A táblázat adatai arra utalnak, hogy a vizsgált mérlegtételek többsége jelentős változást mutat a bázisról a tárgyévra a mintában szereplő vállalkozások esetén, ami számottevő eltérést generál az e mérlegtételeket is tartal-

mazó pénzügyi mutatók értékeire. Ezek az eredmények a vizsgált adathalmaz esetén empirikusan is megalapozottá teszik a tanulmányban megfogalmazott kutatási kérdés vizsgálatának relevanciáját.

Az empirikus vizsgálat

A feltett kutatási kérdések kapcsán végzett empirikus vizsgálatomat logisztikus regresszió alkalmazásával végeztem. A választást a módszer szakirodalmi elterjedtsége, másrészt Zhang – Thomas (2015) azon megállapítása indokolja, miszerint a hitelezői scorecard-ok 95%-a ezzel a módszerrel készül. Az eljárás alkalmazásának előnye, hogy nem követeli meg a magyarázó változók normális eloszlását, viszont érzékeny a kiugró értékek jelenlétére, amely a pénzügyi mutatók egyik alapvető statisztikai sajátossága (Du Jardin, 2010). Ebből adódóan a modellépítést megelőzően fontos szerepet tölt be az adatelőkészítés, melynek célja, hogy a rendelkezésre álló adatbázis statisztikai tulajdonságai minél inkább megfeleljenek az alkalmazni kívánt többváltozós statisztikai eljárás követelményeinek.

Mivel a kiugró értékek azonosítására nincs általánosan elfogadott módszer (Tsai – Cheng, 2012), így erre a célra statisztikai hüvelykujj-szabályokat alkalmaznak a gyakorlatban. Ez a megközelítés véleményem szerint esetleges, mivel előfordulhat, hogy egy adott minta vonatkozásában a kiugró értékek kezelésének egy adott módja hatékony, más adathalmaznál ez azonban nem garantált. A kiugró értékek kezelésére alapvetően kétféle megközelítést alkalmaznak a szakirodalomban: i) a kiugró értékek elhagyása, amely sok esetben a modellezés szempontjából releváns információ elvesztésével járhat (Angelini et al., 2008), ii) a kiugró értékek helyettesítése az adatbázis valamely másik, de már nem kiugró értékével.

Mivel a kiugró értékek jelentősen torzítják a statisztikai modellek illeszkedését, az outlier-ek kezelése elkerülhetetlen. Magam is úgy vélem, hogy a kiugró értékek releváns információt hordoznak, mivel léteznek úgynevezett outlier csődök (Li et al., 2011), melynek lényege, hogy a csődös cégek a működőkhöz képest gyakran minősülnek statisztikai szempontból kiugró értéknek (Lee et al., 2005). A jelen kutatás vonatkozásában további érv a kiugró értékek kezelésére, hogy egyes pénzügyi mutatók nevezője sok esetben nulla volt, melynek leggyakoribb oka, hogy egyes vállalkozások nem rendelkeztek készletekkel, követelésekkel vagy épp árbevétellel a csőd előtti évben. Ilyen esetekben a hányados típusú pénzügyi mutatókat lehetetlen kiszámítani. Az adatvesztés elkerülése érdekében az ilyen esetekben a nulla nevezőt egy egységgel helyettesítettem, ami jelentősen megnövelte az adatbázisban szereplő kiugró értékek számát. A kiugró értékek kezelésének gyakori módja az érintett megfigyelések elhagyása², ez azonban ismételtelen jelentős adatvesztéssel járt volna, ezért az extrém értékek torzító hatásának kiküszöbölésére a változók értékészletének kategorizálását használtam.

A választott megoldás a nélkül teszi lehetővé a kiugró értékek modellben tartását, hogy azokat bármilyen módon helyettesíteni kellene. Ezt egy döntési fát felállító eljárás

alkalmazásával valósítottam meg. A módszertani fejlődésnek köszönhetően számos döntési fa módszer áll rendelkezésre. Ezek közül a választás a CHAID-módszerre esett, melynek lényege, hogy a magyarázó változó értékészletét intervallumokra bontja, majd osztályköz páronként vizsgálja, hogy az osztályközök és az azokban szereplő vállalatok státusa (csődös vagy működő) független-e egymástól. Amennyiben a hipotézist elfogadják, akkor az eljárás a két osztályközt egyesíti. Az eljárás addig folytatódik, amíg csak olyan osztályközök maradnak, amelyek vonatkozásában a statisztikai függetlenség hipotézise elvethető. Az eljárás eredményeképp a magyarázó változó értékészlete osztályközökre bomlik. Az empirikus vizsgálatban minden magyarázó változóra külön alkalmaztam a CHAID-módszert, kategorizálva ezzel azok értékészletét. Ezt követően a logisztikus regressziós modellekben már nem az eredeti pénzügyi mutatókat, hanem azok kategorizált változatát használtam fel; még pontosabban: a regressziók magyarázó változói az egyes pénzügyi mutatók vonatkozásában a CHAID-módszerrel megképzett osztályközök sorszámai voltak.

Hasonló megközelítést alkalmazott a pénzügyi információ tartalmának tömörítésére Kristóf – Virág (2012) is. A szerzők szintén a fentihez hasonló egyváltozós fákat állítottak fel és az azok által a vizsgált vállalkozásokra adott klasszifikációkat, mint dummy változókat használták fel a modellek magyarázó változóiként. Az általam alkalmazott megközelítés ettől csak annyiban tér el, hogy nem az egyes változók alapján adott előrejelzéseket használtam fel új változóként, hanem a CHAID által képzett osztályközök sorszámaikat, ami lehetővé teszi a kiugró értékek modellben tartását, mivel azok valamely szélső osztályközben jelennek meg, viszont az osztályközök sorszáma már nem minősül kiugró értéknek. A változók kategorizálását az SPSS szoftverrel végeztem az alapértelmezett beállítások³ használatával. Ennek megfelelően az eljárás kezdetén a CHAID-módszer minden változónál 10 azonos hosszúságú osztályközből indult ki, melyeket folyamatosan egyesített a korábban bemutatott elvnek megfelelően. Az eljárás eredményeképp kapott osztályközök esetén statisztikailag megalapozottan vethető el, hogy a pénzügyi mutató értéke és a vállalat pénzügyi helyzete független egymástól.

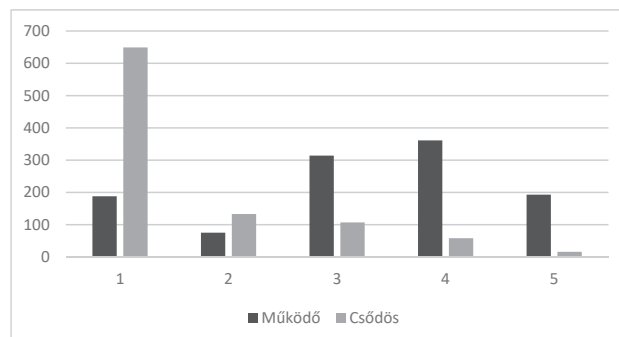
A cash flow/kötelezettségek mutató fentiek szerint kategorizált eloszlását mutatja az 1. és 2. ábra a stock típusú adatok kétféle figyelembevétele mellett.

Mindkét esetben hasonló eloszlást láthatunk, azaz a mutató számítás módja nem gyakorolt jelentős hatást az eloszlásra. Megfigyelhető, hogy a mutató eloszlása a csődös vállalatok esetén erősen balra ferde, ami arra utal, hogy a csődös cégek esetén a mutató értéke jellemzően alacsony. Ezzel szemben a működő cégek eloszlása sajátosnak mondható. Viszonylag sok megfigyelés került ugyanis a csődös vállalatok többségét is tartalmazó első osztályközbe. Ennek az lehet az oka, hogy ezek a vállalatok a mintavétel időpontjában még működőek voltak, de vélhetően pénzügyi problémákkal küzdöttek, amit a relatíve alacsony mutatóérték jelez. Ettől eltekintve a mutató enyhén jobboldali aszimmetriát jelez, ami arra utal, hogy

a működő vállalatokra inkább a magasabb cash flow/kötelezettségek mutató jellemző.

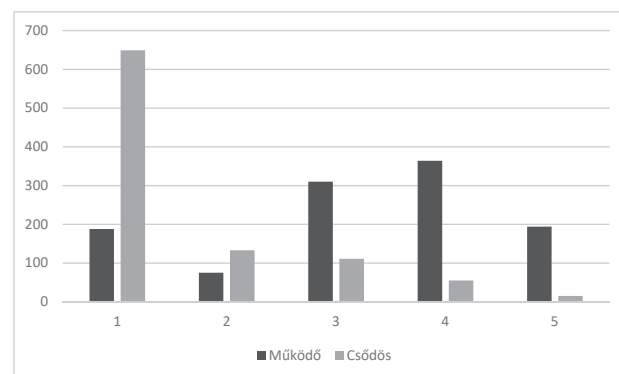
1. ábra

A cash flow/összes kötelezettségmutató eloszlása időszak végi záró értékek használata esetén a CHAID-módszerrel képzett kategóriák alapján



2. ábra

A cash flow/összes kötelezettségmutató eloszlása az állományi adatok átlagolása esetén a CHAID-módszerrel képzett kategóriák alapján



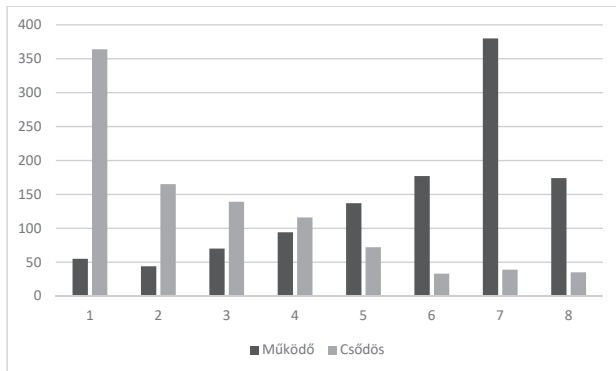
Az eloszlás alakját nem, de a CHAID-módszer alapján megképzett osztályközök számát befolyásolta az állományi adatok figyelembevétele az eszközarányos jövedelmesség mutató esetén. A stock adatok átlagolása esetén a kategóriák száma kevesebb, míg az időszak végi értékek esetén nagyobb volt. A csődös vállalatok esetén a mutató számítás módtól függetlenül erős bal, a működő cégek esetén pedig erős jobboldali aszimmetriát mutat, amint az a 3. és 4. ábrákon látható.

Helytakarékosági okokból csak a fenti két mutató eloszlását mutattam be részletesen. Fontos azonban hangsúlyozni, hogy számos mutató eloszlásában mutatkozik markáns különbség a csődös és működő vállalatok közt. Ezek az egyváltozós tapasztalatok alapot nyújtanak annak feltételezésére, hogy a fent látott eltérések a többváltozós modellek teljesítményében és szerkezetében is tetten érhetők. E vizsgálatok eredményeit ismertetem a továbbiakban.

Az empirikus vizsgálat során a logisztikus regresszió módszerét alkalmaztam, mivel a tudományos kutatásban és a gyakorlati alkalmazásban is ez tekinthető a legnép-

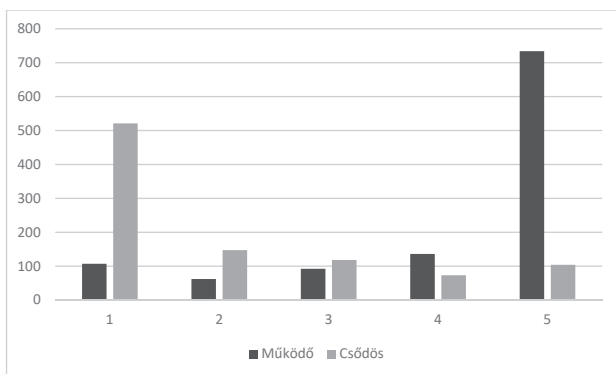
3. ábra

Az eszközarányos jövedelmezőség (ROA) mutató eloszlása időszak végi záró értékek használata esetén a CHAID-módszerrel képzett kategóriákban



4. ábra

Az eszközarányos jövedelmezőség (ROA) mutató eloszlása az állományi adatok átlagolása esetén a CHAID-módszerrel képzett kategóriákban



szerűbb módszertani keretrendszernek. Az elemzés azt vizsgálja, hogy érinti-e a modellek előrejelző képességét, ha az 1. táblázatban vastag betűvel kiemelt változók esetén az átlagolás helyett csak a tárgyévi mérleg záró értékeit vesszük figyelembe. Annak érdekében, hogy a modellek előrejelző teljesítményét mérni tudjunk, szükség van egy úgynevezett modellépítési (tanuló) mintára és egy olyan adatkörre (tesztelő minta), amelyet nem használunk fel a modellépítés során, hanem annak megfigyelésén vizsgáljuk a tanuló mintán felállított modellek teljesítményét. Azonban tekintettel arra, hogy a rendelkezésre álló adatbázis egyszeri véletlenszerű felosztása tanuló és tesztelő mintákra könnyen adhat alapot téves következtetések levonására, ezért a tanulmányban a tízszeres keresztvalidáció módszerét alkalmaztam. Ennek lényege, hogy a teljes adatbázist tíz egyenlő részre osztjuk fel. Ezek közül kilencet használunk fel modellépítésre és a maradék egy tizedet a felállított modell tesztelésére. A módszer lényege, hogy alkalmazása során az adatbázis minden tizede szerepel egyszer a tesztelő minta szerepében. Ez lényegében tíz modell felállítását jelenti, ami lehetővé teszi, hogy a tanulmányból levont következtetések megbízhatóbbak legyenek, ugyanis a modellek előrejelző képességét a tíz

tesztelő mintán kapott eredmény átlagában ítéljük meg, kiküszöbölve az egyszeri véletlenszerű felosztás esetlenségének káros hatásait.

A klasszifikációs módszerek teljesítményét gyakran mérik a találati aránnyal, ami a helyesen besorolt megfigyelések számát viszonyítja az összes megfigyeléshez. A találati arány kiszámításához szükség van egy úgynevezett vágópont meghatározására, amely elválasztja egymástól a csődösnek, illetve működőnek minősítendő megfigyeléseket. Mivel a választás a modellező feladata, így a vágópont értéke befolyásolhatja a találati arányok nagyságát. Ez a probléma azonban nem merül fel abban az esetben, ha a modellek előrejelző teljesítményét a ROC⁴ görbe alatti terület segítségével mérjük, mivel ez a megközelítés az összes lehetséges vágópont mellett vizsgálja a modell teljesítményét. A ROC-görbét egy olyan koordinárendszerben ábrázoljuk, melynek vízszintes tengelyén az egyes lehetséges vágópontok esetén a tévesen csődösnek minősített vállalatok aránya (téves riasztási arány) szerepel, a függőleges tengelyen pedig a helyesen csődösnek minősített vállalatok aránya (helyes riasztási arány). A módszer mögöttes logikája, hogy azok a modellek preferáltak, amelyek adott nagyságú téves riasztás esetén a lehető legnagyobb arányú helyes riasztást adják. Ebből adódóan a minél meredekebben emelkedő ROC-görbék tekinthetők ideálisnak. Mivel a találati arányok értéke minden esetben 0 és 1 közé esik, így a ROC-görbék – ahogy később látható lesz – egy egységnyi területen ábrázolhatók. Minél meredekebb a ROC-görbe, annál nagyobb az alatta lévő terület, ezért a modellek teljesítményének megítélésében általánosan elterjedt mérőszám a ROC-görbe alatti terület használata. Ezeket az értékeket a tíz tanuló és tesztelő minta megfigyelésén egyaránt kiszámítottam.

Az empirikus vizsgálat második lépéseként véletlenszerűen felosztottam az adatbázist tíz egyenlő részre. Ezt követően tíz logisztikus regressziót állítottam fel oly módon, hogy minden esetben az adatbázis egy másik tizede szerepeljen tesztelő mintaként. A magyarázó változók közt először az 1. táblázatban látható pénzügyi mutatók szerepeltek. Ezt követően ugyanazon tíz mintafelosztáson szintén tíz modellt állítottam fel, de ebben az esetben az 1. táblázatban vastaggal kiemelt változók esetén az átlagos értékek helyett a tárgyévi mérleg záró értékeit használtam fel a mutatók kiszámításához. Elsőként vizsgáljuk meg azt, hogy hogyan érinti a modellek felépítését az 1. táblázatban vastag betűvel jelölt mutatók számításmódja. A táblázatban „Modell 1” utal arra az esetre, amikor a kérdéses mutatók számítása a statisztikai irányelveknek megfelelően átlagolással történik, a „Modell 2” kifejezés pedig azon modellekre, ahol az érintett mérlegtételek időszak végi záróérték formájában lettek figyelembe véve. A 3. táblázatban csak azon paraméterek szignifikanciaszintje van feltüntetve, amelyek 5%-os szinten szignifikánsnak bizonyultak.

A tanulmány szempontjából releváns változók közül a cash flow/kötelezettségek mutató csak hat esetben mutatkozott szignifikánsnak a tíz modell közül. Ebből ötször abban az esetben, amikor a mutató számítása során az összes kötelezettség értékét időszak végi záróérték for-

A 10 mintán felállított modellek szerkezete (p érték)

Mutató neve	minta 1		minta 2		minta 3		minta 4		minta 5		minta 5		minta 7		minta 8		minta 9		minta 10	
	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2
Likviditási ráta	0 ⁵	0	0	0,001	0	0	0	0	0	0,005	0	0,001	0	0,001	0	0	0	0	0	0
Likviditási gyorsráta	-	-	0,001	-	0,015	0,024	0,01	-	-	-	0,048	-	0,021	-	0,032	-	0,007	-	0,002	-
Cash flow/Kötelezettségek	-	-	-	0,006	-	-	-	0,007	-	-	-	0,001	-	0,001	-	0,004	-	-	0,015	-
Cash flow/Rövid lejáratú kötelezettségek	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tőkeellátottság	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Eszközök forgási sebessége	-	-	0	-	0,002	-	0	-	-	0,004	-	0,01	0	-	-	0,025	-	-	0	-
Készletek forgási sebessége	0	-	0	-	0	0	0	0,001	0	-	0	0,034	0	-	0	-	0	0,024	0,001	-
Követelések forgási ideje	0	0	0,027	0	0,003	0,006	0,05	0	0	0	0	0,04	0,004	0	0	0	0	0	-	0
Eladósodottság	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Saját tőke aránya	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bonitás	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Arbevétel-arányos nveresség	-	-	-	-	-	0,03	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Eszközarányos nveresség	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,005
Követelések/Rövid lejáratú kötelezettségek	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nettó forgótőke aránya	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Vállalat mérete (1)	0,015	-	0,034	0,017	0,031	-	0,006	0,01	-	-	0,013	0,001	0,003	0,004	-	0,002	-	-	0,001	0,002
Eredménytartalék aránya	0	0	0,003	0,005	0	0	0	0	0	0	0	0	0,002	0,004	0	0	0,003	0,003	0	0
Pénzeszközök aránya	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vállalat mérete (2)	-	0,003	-	-	-	-	-	-	-	-	-	-	0,036	-	-	-	-	-	-	-

májában vettem figyelembe. Csak egyetlen esetben volt szignifikáns átlagolás esetén. Megfigyelhető az is, hogy ha az egyik esetben (például átlagolás nélkül) szignifikáns, akkor a másik esetben (átlagolással) nem, és fordítva. A rendelkezésre álló minta esetén tehát e változó esetén célszerűbbnek tűnik az időszak végi záróérték használata az átlagolás helyett. Érdekes eredmény, hogy a cash flow/rövid lejáratú kötelezettségek mutató számításától függetlenül egyetlen esetben sem volt szignifikáns.

Az eszközök forgási sebessége nyolc esetben volt szignifikáns: ebből öt esetben a számítás átlagolással történt, három esetben nem. Ebben az esetben is igaz, hogy a számítás mód kizáró erővel bír: azaz, ha például átlagolás nélkül szignifikáns volt a mutató, akkor átlagolással nem, és fordítva. E változó esetén a tanulmányhoz felhasznált adatbázis esetén a mérlegtételek átlagolása tűnik praktikusabb választásnak.

A készletek forgási sebessége átlagolás esetén minden esetben szignifikáns mutató volt, ugyanakkor volt négy olyan minta is, ahol az időszak végi záróérték használata mellett is szignifikáns volt a mutató. Azaz ebben az esetben ismételt az átlagolás tekinthető preferáltabb eljárásnak az elemzett vállalatok vonatkozásában.

A követelések forgási ideje szinte minden esetben szignifikáns volt számításától függetlenül. Egyetlen kivételt az utolsó minta képez, ahol átlagolással a mutató nem minősült szignifikánsnak. E mutató tehát úgy tűnik, számításától függetlenül szignifikáns magyarázó erővel bír a mintában szereplő vállalkozások jövőbeli csődjére vonatkozóan. Ugyanez mondható el – de ebben az esetben kivétel nélkül – az eszközarányos jövedelmezőség mutatójáról is.

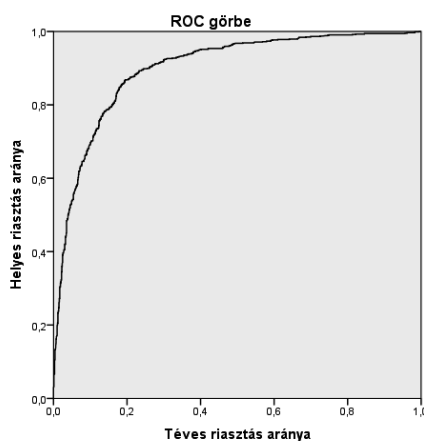
Amint a 3. táblázat adataiból látható, a modellek szerkezetét csak kis mértékben érintette a vizsgálat szempontjából releváns mutatók számítás módjának változása. Kérdés azonban, hogy a kisebb szerkezeti eltérések a modellek illeszkedését és előrejelző teljesítményét mennyiben érintik. A fenti elemzések kapcsán felmerülhet annak lehetősége is, hogy az állományi adatok figyelembevételét a korábbi eredmények alapján valósítsuk meg. Azaz a cash flow/kötelezettségek, valamint a követelések forgási ideje esetén a mérlegtételek az időszak végi záró érték formájában, a többi mutató esetén pedig átlagolva, mivel a 10 mintafelosztás esetén ezekben az esetekben voltak az egyes mutatók a legtöbbször szignifikánsak. Két kivételt a cash flow/rövid lejáratú kötelezettségek, illetve az esz-

közarányos jövedelmezőség mutatója jelentett. Előbbi semmilyen formában nem volt szignifikáns egyszer sem, utóbbi pedig minden esetben szignifikáns volt. Ezekben az esetben az átlagolás került alkalmazásra. Az így felállított modellek eredményei a „Modell 3” címszó alatt jelennek meg a későbbiekben.

A bemutatott három megközelítés vonatkozásában a tíz mintafelosztás mindegyikére kiszámítottam a modellek ROC-görbe alatti területét. Ezeket az első tanuló-tesztelő felosztás esetén az 5. és 6. ábrák szemléltetik. Mivel a görbék alakja a többi felosztás esetén is hasonló, ezért helytakarékosági okokból csak az első felosztás esetén kapott görbéket szemléltetem.

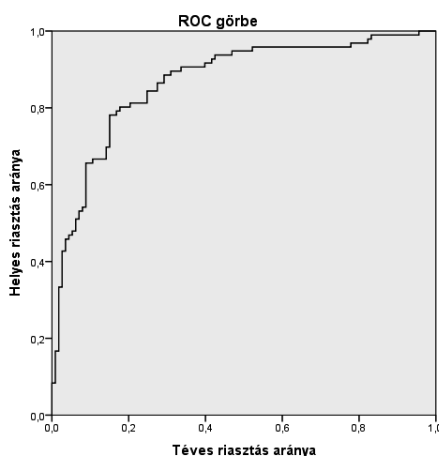
5. ábra

Az első tanuló mintán felállított modell ROC-görbéje



6. ábra

Az első tanuló mintán felállított modell ROC-görbéje az első tesztelő minta megfigyelésein



A tíz mintán felállított modellek ROC-görbe alatti területeit mutatja a 4. táblázat.

A táblázat eredményei azt mutatják, hogy az 1-es modell (a kérdéses mutatók átlagolással számítva) minden esetben jobb illeszkedéssel és magasabb előrejelző teljesítménnyel jellemezhető a tíz mintafelosztás mindegyikén,

mint a 2-es modell (a kérdéses mutatók esetén a mérlegadatok időszak végi záróérték formájában szerepeltetve). A különbség mértéke 0,1 és egy százalékpont közt mozgott, ami átlagosan 0,79 százalékponttal magasabb előrejelző teljesítményt és 0,59 százalékponttal jobb illeszkedést eredményezett a tíz felosztás átlagában (azonos szóródás mellett). A 3-as modell – ahol a kérdéses mutatókat a szerint számítottam, hogy mikor mutatták a legjobb teljesítményt a tíz felosztás többségében – hasonló eredményeket hozott, mint az 1-es modell, ebből adódóan használata elvethető, ugyanis nincs érv az alkalmazása mellett: a mutatók számításmódja nem egységes benne és nem is nyújt jobb eredményt ahhoz képest, amikor a kérdéses mutatók számítása egységesen átlagolással történt.

4. táblázat

A tízszeres keresztvalidáció eredményei (ROC-görbe alatti terület)

Minta	Modell 1		Modell 2		Modell 3	
	Tanuló	Teszt	Tanuló	Teszt	Tanuló	Teszt
1. minta	0,900	0,929	0,893	0,925	0,899	0,926
2. minta	0,901	0,914	0,895	0,897	0,901	0,912
3. minta	0,903	0,903	0,895	0,896	0,902	0,906
4. minta	0,906	0,881	0,899	0,874	0,905	0,877
5. minta	0,904	0,889	0,899	0,883	0,904	0,885
6. minta	0,908	0,854	0,903	0,845	0,907	0,854
7. minta	0,901	0,931	0,891	0,925	0,899	0,935
8. minta	0,897	0,914	0,895	0,904	0,897	0,909
9. minta	0,901	0,874	0,898	0,872	0,902	0,879
10. minta	0,905	0,879	0,899	0,869	0,905	0,877
Átlag	0,903	0,897	0,897	0,889	0,902	0,896
Szórás	0,003	0,024	0,003	0,024	0,003	0,024

Kérdésként merülhet fel az Olvasóban, hogy a modellezést megelőző kategorizálás alkalmazása mennyiben érintette a modellek teljesítményét. Korábban utaltam arra, hogy ezt a kiugró értékek torzító hatásának kiküszöbölése teszi szükségessé. Ennek hatékonyságát úgy vizsgáltam, hogy a fenti elemzést elvégeztem a korábban részletezett kategorizálás nélkül is. Helytakarékosági okokból a számítások részletes eredményeit nem közlöm, csak a legfontosabb tapasztalatokat emelem ki: a kiugró értékek generálta jelentős szóródás miatt alig volt olyan eset, amikor a vizsgálat szempontjából releváns pénzügyi mutatók közül bármelyik szignifikáns lett volna. Ebből adódóan a mindkét modell ROC-görbe alatti területeinek átlaga megegyezett a 10 mintán (0,774). A tesztelő mintákon az 1-es modell átlagos értéke 0,752, a 2-es mintán pedig 0,746. Összevetve a korábbi eredményekkel: ez a nagyságrendileg 14 százalékpontos differencia egyértelműen rámutat a kiugró értékek torzító hatására a logisztikus regresszió esetén; továbbá alátámasztja az e célra alkalmazott kategorizálás hatásosságát.

A kapott eredmények röviden összefoglalva úgy értékelhetők, hogy az átlagolás elmulasztása az érintett mutatóknál egy kismértékű teljesítményromlással párosul. Ugyan a bemutatott eltérések közül a 0,1 százalékpontos differencia a gyakorlati hitelkockázati modellezésben is elhanyagolható, ugyanakkor az egy százalékpontos eltérés már pénzben kifejezve is mérhető lehet egy hitelezési döntésekkel foglalkozó vezető számára. Ez a marginális teljesítménytöbblet elérhető a pénzügyi mutatók számítása során némi többletmunkával, illetve némileg nagyobb adatigénnyel (az átlagoláshoz a tárgyév és a bázisév adataira is szükség van). Az elméleti kutatások szempontjából a bemutatott eredmények arra utalnak, hogy az átlagolás elmulasztása a vizsgált mutatóknál nem torzítja jelentősen a modellek teljesítményét, így ez a „hanyagság” nagy valószínűséggel nem okozza a modellek teljesítményének jelentős romlását. A gyakorlati döntéshozók számára azonban az eltérés számottevő lehet, így az utóbbi esetben célszerű a pénzügyi mutatók számításánál a tankönyvek által javasolt formulákat precízen követni az előrejelző teljesítmény maximalizálása érdekében.

Összegzés

A statisztikai irányelvek alapján abban az esetben, amikor egy hánados stock és flow típusú adatokat vet össze, akkor a stock adatok esetén az időszak nyitó és záró értékének átlagát szükséges figyelembe a számítás során. A tanulmány kiindulópontja az a megfigyelés volt, hogy számos olyan publikáció van a csődelőrejelzés területén a hazai és a nemzetközi szakirodalomban egyaránt, ahol a szerzők vagy nem adják meg pontosan, hogy hogyan számították az ilyen jellegű mutatókat, vagy nem végzik el ezt az átlagolást. A cikk azt a hiány kívánta pótolni, hogy az átlagolás elmulasztásának hatására vonatkozóan még nem publikáltak empirikus vizsgálatot. Ezt erősíti az is, hogy egyetlen olyan publikációval sem találkoztam még, ahol a szerzők megindokolták volna, hogy miért nem végzik el az átlagolást. Ebből adódóan a tanulmány arra vállalkozott, hogy empirikus kutatást végezzen annak kiderítésére, hogy az átlagolás elmulasztása esetén a logisztikus regresszióra épülő csődelőrejelző modellek előrejelző teljesítménye változik-e ahhoz képest, amikor a mutatók számítása során követjük a korábban ismertetett statisztikai elveket.

A feltett kutatási kérdés vizsgálata céljából saját adatgyűjtésből származó mintát használtam. A minta elemei Magyarországon működő vállalkozások, melyek körében a fizetéseképtelenséget jogi értelemben definiáltam annak alapján, hogy az érintett vállalkozással szemben sor került-e felszámolási vagy csődeljárás megindítására. E definíció alkalmazását az indokolta, hogy az adatokat nyilvánosan hozzáférhető adatbázisokból kellett manuálisan összegyűjteni, ahol ez volt az egyetlen elérhető információ a vállalatok fizetéseképtelenségére vonatkozóan. A minta 2094 megfigyelésből állt, melyek közt 56%-ot tettek ki a működő vállalkozások, 44%-ot pedig a csődösek. A vállalatok pénzügyi helyzetét a hazai szakirodalomban leggyakrabban alkalmazott pénzügyi mutatók köréből válogattam,

kiegészítve néhány saját megfontolás alapján választott mutatóval. A 19 mutató közül hat volt olyan, amely stock és flow típusú adatokat viszonyított egymáshoz.

Az empirikus vizsgálatot a szakirodalomban és a hitelkockázati modellezés gyakorlatában leggyakrabban alkalmazott módszertannal, a logisztikus regresszió alkalmazásával végeztem. Mivel az eljárás érzékeny a kiugró értékek jelenlétére, azok torzító hatását egyváltozós CHAID-alapú döntési fák segítségével küszöböltem ki. A döntési fák eredményeképp kapott kategorizált pénzügyi mutatókat használtam fel a regressziós modellek magyarázó változói közt. A modellek előrejelző képességét tízszeres keresztvalidáció alkalmazásával becsültem meg. A vizsgálat során kétféle modellt állítottam fel: az egyikben a stock és flow adatokat összevető pénzügyi mutatókat a statisztikai elveknek megfelelően számítottam ki, míg a másik esetben ezeket a változókat úgy kalkuláltam, hogy a stock típusú adatok esetén a nyitó és záró érték átlaga helyett csak a tárgyévi záró értéket használtam fel.

A modellek teljesítményét a szakirodalomban és a gyakorlatban egyaránt általánosan elterjedt ROC-görbe alatti területtel mértem. A tíz mintafelosztás mindegyikében kis mértékben jobb teljesítményt mutattak azok a modellek melyek magyarázó változói közt olyan mutatók szerepeltek, ahol a számítás a statisztikai elveknek megfelelően történt. Ki kell emelni ugyanakkor, hogy a különbség több esetben csak marginális és a tíz mintafelosztás átlagában nem érte el az egy százalékpontot. Ebből adódóan a vizsgált adatbázis vonatkozásában kijelenthető, hogy az átlagolás elmulasztása a megfelelő mutatóknál nem eredményezte az előrejelző képesség jelentős romlását. Ez magyarázattal szolgál arra, hogy miért gyakori eset, hogy a témában megjelent publikációk szerzői miért nem végzik el ezt az átlagolást, vagy miért nem részletezik annak tényét, hogy ezt megtették-e. A tanulmány e gyakorlat mögé keresett empirikus magyarázatot, amelyre vonatkozóan hasonló kutatással eddig még nem találkoztam. A gyakorlati modellezés szempontjából a dokumentált különbség azonban már számottevő lehet, így a tanulmány fő következtetése a gyakorlati alkalmazás szempontjából, hogy a pénzügyi mutatókat is tartalmazó csődelőrejelző modellek esetén célszerű az állományi és folyamat szemléletű adatokat összevető hánadosok számítása során a stock adatokat átlagolni az előrejelző teljesítmény növelése érdekében.

A bemutatott elemzés eredményei mellett szót kell ejteni annak korlátairól is. Az empirikus vizsgálatához szükséges adatokat manuális adatgyűjtés keretében gyűjtöttem össze nyilvános adatbázisokból, ahol a vállalatok fizetéseképtelenségére vonatkozóan az egyetlen rendelkezésre álló információ az, hogy azokkal szemben sor került-e felszámolási vagy csődeljárás megindítására. Kérdés lehet azonban, hogy a gyakorlati hitelkockázati modellezésnél relevánsabbnak tekinthető 90 napot meghaladó késedelembe esés vonatkozásában is érvényesek-e az empirikus vizsgálatok alapján levont következtetések?

A bemutatott elemzés másik korlátja abból fakad, hogy a nyilvános adatbázisokban csak a név és cégjegyzékszám alapján van lehetőség keresésre, ebből adódóan az empi-

rikus vizsgálathoz felhasznált minta heterogén volt a vállalatok kora, mérete, tevékenységi köre, stb. tekintetében. Ebből adódóan pedig kérdéses lehet, hogy levont következtetések milyen mértékben érvényesek homogénebb megfigyelések vonatkozásában. E kérdésekre jövőbeli kutatások adhatnak majd választ.

Jegyzet

- 1 http://e-beszamolo.im.gov.hu/oldal/beszamolo_kereses
- 2 Gyakran alkalmazott megközelítés még az értékészlet csonkolása az , illetve valamely szélső percentilis mentén.
- 3 A CHAID-módszernél ez a következőket jelentette: a fák maximális mélysége: 3 szint; a szülő ágak minimális elemszáma 100, a gyerek ágak minimális elemszáma 50 megfigyelés; a kategóriák egyesítésének és felosztásánál használt szignifikancia szint 5%, a szignifikancia értékek korrekciójára a Bonferroni módszer szerint került sor; a függetlenség-vizsgálatok során a Pearson-féle KHI-négyzet statisztikát használtam; az iterációk maximális száma 100 volt; a minimális változás a várható gyakoriságra az egyes celláknál 0,001. A változók értékészlete tíz egyenlő osztályközre lett felosztva az eljárás kezdetén. Az alapbeállítások használata a logisztikus regresszióval a következőket jelentette: A változók beléptetéséhez használt szignifikancia szint 5%, a kiléptetésnél 10%; az iterációk maximális száma 20. A modellépítés során a változók a likelihood arányon alapuló stepwise módszer szerint lettek szelektálva.
- 4 A rövidítés a receiver operating characteristic angol kifejezésből ered.
- 5 A táblázat 0 értékei esetén az empirikus szignifikanciaszint 0,000, míg a „-”, arra utal, hogy a mutató nem volt 5 %-os szinten szignifikáns az adott modellben.

Felhasznált irodalom

- Amendola, A. – Restaino, M. – Sensini, L.* (2011): Variable selection in default risk models. *The Journal of Risk Model Validation*, Vol. 5., No. 1., Spring, p. 3-19.
- Angelini, E. – Di Tollo, G. – Roli, A.* (2008): A neural network approach for credit risk evaluation. *The Quarterly Review Economics and Finance*, Vol. 48., No. 4., p. 733-755.
- Bozsik, J.* (2010): Artificial Neural Networks in Default Forecast. *Proceedings of the 8th International Conference on Applied Informatics*, Eger, Hungary, January 27-30, 2010, Vol. 1. p. 31-39.
- Chen, J. H.* (2012): Developing SFNN models to predict financial distress of construction companies. *Expert Systems with Applications*, Vol. 39., No. 1., p. 823-827.
- Du Jardin, P.* (2010): Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*, Vol. 73., No. 10-12., p. 2047-2060.
- Emel, A. B. – Oral, M. – Reisman, A. – Yolalan, R.* (2003): A credit scoring approach for the commercial banking sector. *Socio-Economic Planning Sciences*, Vol., 37., No. 2., p. 103-123.
- Fantazzini, D. – Figini, S.* (2009): Random survival forests models for SME credit risk measurement. *Methodology and Computing in Applied Probability*, Vol. 11., No. 1., p. 29-45.
- Horta, I. M. – Camanho, A. S.* (2013): Company failure prediction in the construction industry. *Expert Systems with Applications*, Vol. 40., No. 16., p. 6253-6257.
- Hunyadi László – Vita László* (2004): *Statisztika közgazdászoknak*. Budapest: Központi Statisztikai Hivatal
- Ic, Y. T. – Yurdakul, M.* (2010): Development of a quick credibility scoring decision support system using fuzzy TOPSIS. *Expert Systems with Applications*, Vol. 37., No. 1., p. 567-574.
- Kim, M. J. – Han, I.* (2003): The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. *Expert Systems with Applications*, Vol. 25., No. 4., p. 637-646.
- Kristóf, T. – Virág, M.* (2012): Data reduction and univariate splitting – Do they together provide better corporate bankruptcy prediction? *Acta Oeconomica*, Vol. 62., No. 2., p. 205-227.
- Lee, T. S. – Chiu, C. C. – Lu, C. J. – Chen, I. F.* (2002): Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, Vol. 23., No. 3., p. 245-254.
- Li, H. – Adeli, H. – Sun, J. – Han, J. G.* (2011): Hybridizing principles of TOPSIS with case-based reasoning for business failure prediction. *Computers & Operations Research*, Vol. 38., No. 2., p. 409-419.
- Lin, F. – Liang, D. – Yeh, C. C. – Huang, J. C.* (2014): Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*, Vol. 41., No. 5., p. 2472-2483.
- Min, J. H. – Jeong, C.* (2009): A binary classification method for bankruptcy prediction. *Expert Systems with Applications*, Vol. 36., No. 3., Part 1., p. 5256-5263.
- Sánchez-Lasheras, F. – De Andrés, J. – Lorca, P. – De Cos Juez, F. J.* (2012): A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications*, Vol. 39., No. 8., p. 7512-7523.
- Santos Silva, J. M. C. – Murteira, J. M. R.* (2009): Estimation of default probabilities using incomplete contracts data. *Journal of Empirical Finance*, Vol. 16., No. 3., p. 457-465.
- Sun, L.* (2007): A re-evaluation of auditors' opinions versus statistical models in bankruptcy prediction. *Review of Quantitative Finance and Accounting*, Vol. 28., No. 1., p. 55-78.
- Tsai, C. F. – Cheng, K. C.* (2012): Simple instance selection for bankruptcy prediction. *Knowledge-Based Systems*, Vol. 27., p. 333-342.
- Virág Miklós – Kristóf Tamás – Fiáth Attila – Varsányi Judit* (2013): *Pénzügyi elemzés, csődelőrejelzés, válságkezelés*. Budapest: Kossuth Kiadó
- Virág Miklós – Kristóf Tamás* (2005): *Az első hazai csődmodell újraszámítása neurális hálók segítségével*. *Közgazdasági Szemle*, 52. évf., 2. szám, p. 144-162.
- Virág Miklós – Kristóf Tamás* (2006): Iparági rátákon alapuló csődelőrejelzés sokváltozós statisztikai módszerekkel. *Vezetéstudomány*, 37. évf., 1. szám, p. 25-35.
- Virág Miklós – Kristóf Tamás* (2006): Többdimenziós skálázás a csődelőrejelzésben. *Vezetéstudomány*, 37. évf., 1. szám, p. 25-35.
- Virág, M. – Nyitrai, T.* (2013): Application of support vector machines on the basis of the first Hungarian bankruptcy model. *Society and Economy*. Vol. 35., No. 2., p. 227-248.
- Virág, M. – Nyitrai, T.* (2014): Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model. *Acta Oeconomica*, Vol. 64., No. 4., p. 419-440.
- Zhang, J. – Thomas, L. C.* (2015): The effect of introducing economic variables into credit scorecards: an example from invoice discounting. *Journal of Risk Model Validation*. Vol. 9., No. 1., p. 57-78.
- Zhou, L.* (2013): Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, Vol. 41, p. 16-25.