

DR. ARATÓ MIKLÓS

Biztosítások területi kockázatának meghatározásáról

A biztosítóknál dolgozó aktuáriusok egyik legfontosabb feladata hagyományosan a díjszámítás. A szükséges díjszintet nagyon sok tényező együttesen határozza meg. Ilyenek például a biztosítási szerződés azon jellemzői, amelyek befolyásolják a károk számát és nagyságát, a törléseket. Igen fontos tényezők még a költségek, a jutalékok, a tőkeköltségek és a tulajdonosok nyereségelvárása is.

Jelen cikkben a szerződők lakcímének, telephelyének a károk számára és nagyságaira történő hatásával foglalkozom. Ezen hatások helyes meghatározása segítheti a biztosítót a veszteségek elkerülésében, az állományszerzésben, valamint az állományvesztés elkerülésében. A biztosítási gyakorlatban a területi számítások gyakran ad hoc történnek. Munkámban eligazítást kívánok adni a megfelelő modell kiválasztásához. Ehhez egyrészt bemutatom az alkalmazható modellek körét, másrészt ismertetem a modellkiválasztást segítő módszereket. Az egyetemi oktatás keretén belül ezekre a témákra – sok más fontos terület mellett – nem jut elegendő idő, és a téma specialitása miatt nem is tartom feltétlenül ott oktatódnak. Az ismertetésre kerülő eljárásokra, módszerekre vonatkozóan magyar nyelvű irodalom nem áll rendelkezésre, így tanulmányomat e szempontból is hiánypótlásra számom. Reményeim szerint a következőkben ismertetett eljárások más alkalmazási területeken tevékenykedők számára is hasznosíthatók.

Az 1. fejezetben bemutatom a magyarországi tapasztalatokat a területi díjdifferenciálás témaköréből. Történeti áttekintés után a díjak területi megkülönböztetésének jelenlegi helyzetét vizsgálom meg. A területi díjkülönbözőségeken túl azt is bemutatom, milyen tényezők befolyásolhatják ezeket. Külön kitérek a katasztrófa károk különbözőségére, mivel meggyőződésem szerint ezek becslésére nem elegendő csak a saját kártapasztalatok figyelembevétele, hanem hidrológiai, geofizikai modellek kiépítésére is szükség van.

A területi epidemiológia (*spatial epidemiology*) és a betegségtérképezés (*disease mapping*) a területi, térbeli statisztika (*spatial statistics*) egyik leginkább kidolgozott területe. Ezen elemzések során – amennyiben a vizsgálatokba sok területi egységet vonnak be – egy-egy egységben a megfigyelések (megbetegedések) száma általában alacsony, ezért ahhoz, hogy a területi kockázatokat megbecsüljék, szükség van a háttérváltozókra összefüggőségi struktúrákat feltételezni. Ez lényegében bayesi megközelítésnek felel meg. Az alkalmazott modelleknél általános, hogy szomszédsági feltételeket és szomszédsági összefüggéseket állítanak fel. Igen gyakori a Markov-tulajdonság feltételezése is, tehát, hogy egy területi egység kockázata az összes többi kockázattól csak a szomszédjai kockázatán keresztül függ. A legnagyobb hatásúak talán Clayton, Kaldor (1987) és Besag, York, Mollié (1991) munkái voltak. Ezek a cikkek inspirálták az első, területi kockázatokkal foglalkozó biztosításmatematikai cikkeket is. Az általam legfontosabbnak tartott területi aktuáriusi modelleket a második fejezetben mutatom be.

A harmadik fejezetben egy magyarországi nem életbiztosítási állomány területi szempontú elemzését mutatom be. Az ehhez kapcsolódó adatbázis-állomány több mint félmillió

szerződés és több mint százezer kár adatát tartalmazza. Az adatokra több modellt is illesztettünk. Az illesztés mellett bemutatom azt is, hogy milyen módszerekkel választhatunk a különböző lehetőségek közül. Kiemelten fontosnak tartom az egyedi szerződés- és káradatok felhasználását. Ez sokkal megbízhatóbbá teszi a választási döntést, mint a publikációk döntő részében felhasznált adataggregálás. Annak ellenére, hogy üzleti titkok megsértése nélkül csak az adatok egy részét mutathatom be, remélem, hogy az aktuáriusok fel tudják használni munkájukban az elemzés módszereit.

Magyarországi tapasztalatok

Magyarországon a II. világháború előtt igen fejlett biztosítási piac működött. Ennek megfelelő volt a biztosításmatematika színvonala is, de a nemzetközi gyakorlatnak megfelelően a biztosításmatematikások (aktuáriusok) elsősorban az életbiztosításokkal foglalkoztak. A nem életbiztosítások díjainál már a 19. században is megjelent vidéki, nem vidéki megkülönböztetés. Nincs információ arról, hogy milyen díjszámítási módszereket alkalmaztak a díjmegkülönböztések meghatározásához.

A szocializmus korszakában 1952-től 1986-ig egyetlen biztosító működött: az Állami Biztosító. Ebben az időszakban a nyereséges működés kis szerepet játszott és a nem életbiztosítási díjszámításra csekély figyelmet fordítottak. 1986-ban a vállalat két részre bomlott, az Állami Biztosítóra és a Hungária Biztosítóra. 1988-tól kezdtek megjelenni további versenytársak is a magyar piacon, napjainkban már néhány tucat biztosító működik. Ez a változás hozzájárult a kalkulációs eszközök fejlesztéséhez.

1986-ban létrehozták az Állami Biztosításfelügyeletet, amelynek korai korszakában egyik legfontosabb feladata a termékengedélyezés volt. A díjkalkulációk tekintetében az engedélyezés 1991-től már nem a díjak szintjére koncentrált, hanem az alkalmazott módszerek megfelelőségét vizsgálták. A termékengedélyezés először a nem életbiztosításoknál szűnt meg az 1990-es évek közepén, utána nem sokkal az életbiztosításoknál is. A kötelező gépjármű-felelősségbiztosításnál (KGFB) azonban még a 2000-es évek elején is csak a felügyelet által engedélyezett díjakat lehetett alkalmazni.

Az új cégek 1988-as megjelenésétől kezdve a biztosítók fokozatosan egyre több szempontot kezdtek figyelembe venni a kockázatok díjának meghatározásánál.

Az életbiztosítások esetében – a hagyományoknak megfelelően – elsődlegesen az életkor és a nem (ez utóbbi megkülönböztetés uniós előírás miatt már nem lehetséges) határozta meg a díjat, továbbá betegség és néha dohányzás és elhízás miatt emelték meg az alapidíjat. Más tényezőket sem korábban sem most nem vesznek figyelembe.

A nem életbiztosításoknál a szerződés jellege szerint igen sok tényezővel számolnak. Ilyenek például a gépjármű-biztosításoknál a gépkocsi típusa, színe, teljesítménye, a lakás- és egyéb ingatlan biztosításoknál a lakóépület magassága, tetőzete, építési éve, az egyéb kockázatoknál akár a szerződő életkora, a szerződés megkötésének éve. A budapesti – nem budapesti megkülönböztetés már az 1990-es évek elejétől megjelent a casco-, a lakás- és a balesetbiztosításoknál. Többen a vidéki nagyobb városokra is külön díjtételeket határoztak meg. Nagyon részletes területi díjstruktúrát azonban most sem alkalmaznak a kötelező gépjármű-felelősségbiztosítás (KGFB) kivételével.

A KGFB biztosítási alapon (nem benzinárba építve) 1991 júliusában indult. Több éven keresztül a díjat központilag határozták meg és az a személygépkocsiknál csak a gépkocsi

motorjának térfogatától függött. A díjliberalizáció után a biztosítótársaságok egyre több területi kategóriát is bevezettek.

A biztosítók jogszabályi kötelezettsége, hogy KGFB díjtételeiket nyilvánossá tegyék. A <http://www.mabisz.hu/hu/kgfb-dijtarifak.html> oldalon valamennyi olyan biztosító díja megtalálható, amelyek KGFB-vel foglalkoznak. Ez lehetővé teszi, hogy összehasonlítsuk a különböző cégek területi besorolásainak hasonlóságait és különbözőségeit. Az összehasonlíthatóság lehetősége egyedülálló, egy másik terméknel sincs mód ilyen egyszerűen megfigyelni a biztosítók tarifáit.

Az egyes tarifákban a területi csoportok száma 5 és 40 közötti. A legdrágább és legolcsóbb díj terület miatti eltérése is biztosítónként más és más, leggyakrabban az arány kétszeres.

A tarifákban a területi besorolás miatt elég jelentősek az eltérések az egyes biztosítók között. Ez azonban nem feltétlenül jelenti azt, hogy a végső, megállapított díjakban is jelentősek az eltérések, hiszen a díjmeghatározásánál a terület mellett sok más tényezőt is figyelembe vesznek. Az egyéb kockázati szempontoknál (kor, nem, gépkocsi típusa stb.) alkalmazott eltérő szorzók összességében hasonló díjat eredményezhetnek.

Mitől függnek a területi díjeltérések? Az alábbiakban felsorolunk néhány lehetséges tényezőt.

- Várható kárkifizetés. A díjmeghatározásnál ez a legfontosabb szempont.
- Szerződések várható élettartama, törlési valószínűségek.
- Más termékek eladási lehetősége.
- Kockázatok koncentrációja. A 2016-tól érvényes az Európai Unióban a Szolvencia 2 szabályozás sokkal nagyobb hangsúlyt fektet a kockázatoktól függő szavatolótőke-szükségletre. A katasztrófakockázat miatti tőkeszükséglet erősen függ attól, hogy közeli területeken mennyi a lakás- és a vállalativagyron-biztosítási összeg. A szavatolótőke-szükséglet most már egyre több biztosítónál díjbefolyásoló tényező és a katasztrófakockázat miatt főleg Budapesten növelheti a díjat.
- Más biztosítók díjainak hatása. A verseny miatt a biztosítók eltéríthetik díjaikat az általuk reálisnak tartottól is.

A díjak szintjét még erősen befolyásolják a költségek, ezen belül a jutalékok. Ezeket azonban Magyarországon nem szokták területi alapon differenciálni.

Cikkünkben csak a kárkifizetések területi különbözőségének vizsgálatával fogunk foglalkozni.

Milyen adatok felhasználásával történnek a számítások? Amennyiben a biztosító most kezdi a működését, vagy teljesen új termék terjesztését kezdi meg, akkor leginkább más biztosítók díjtételeire és a viszontbiztosítók tanácsaira alapoz. Élet-, betegség- és baleset-biztosítások esetében a saját tapasztalattal rendelkező biztosítók is fel szokták használni az elérhető KSH, ONYF és más statisztikákat. A halandósági eltérések kistérségi szintű eltéréseivel kapcsolatban megemlíthető Arató–Dryden–Taylor (2006) és Bálint (2011) tanulmányai. Azonban az egész népességre vonatkozó mortalitási, egészségügyi adatok csak korlátozottan alkalmasak biztosítói felhasználásra. Például az üzleti biztosítással rendelkezők halandósága 30–50%-kal kisebb a néphalandóságnál. Ezt a területi differenciálásnál is figyelembe kell venni, mert a halálozásokat nagyban befolyásoló kereseti, foglalkoztatási

adatok területi különbségei sokkal kisebbek a biztosítottak csoportjára szűkítve. Casco biztosításoknál fel szokták használni az autólopási és közlekedési baleseti adatokat, a lakás-biztosítások betörési lopás moduljánál pedig a betörések számát.

Nem katasztrófakockázatoknál a legfontosabb forrás a saját szerződés- és káradatbázis. Ezek felhasználhatóságánál kritikus a biztosító adatnyilvántartás-vezetésének megfelelő-sége. A területi adatok vonatkozásában például szerencsés, ha nemcsak a szerződő lakhe-lyének címe, hanem a kockázatviselés helyének címe is rendelkezésre áll. Gyakran csak a postai irányítószám van feldolgozható formában rögzítve, ez azonban több esetben még a települést sem azonosítja. A károknál alapvető, hogy a kár időpontja, a bejelentés időpontja és a kifizetés(ek) időpontja is rögzítve legyen. A többi nem életaktuáriusi számításhoz ha-sonlóan mindig tekintetbe kell venni a bejelentési és kifizetési késlekedést. Elvileg előfor-dulhat, hogy ebben is különbség van a különböző területek között.

A természetikatasztrófa-kockázatok közül a viharokára a biztosítók jelentős része megfelelő saját tapasztalattal rendelkezik, bár az is kérdéses, hogy Magyarországon a vi-harok tekinthetők-e katasztrófakockázatnak. Az árvíz- és földrengéskockázat azonban semmi esetre sem elemezhető csak a saját adatok segítségével. Szerencsére az utolsó iga-zán jelentős károkat okozó árvíz 1970-ben volt, míg az utolsó jelentős földrengés a 16. században fordult elő Pest-Budán. Ennek ellenére az ilyen károk egyáltalán nem elhanya-golthatók. Gondoljunk bele, ha egy 30 millió forintos ház helyét átlagosan 100 évente önti el az árvíz, akkor az ilyen esetre 15 millió forintos kárt feltételezve, a várható éves kárki-fizetés csak árvízkárokra 150 ezer forint. Minden költség, kockázati ráhagyás nélkül! Je-lenleg a minden kockázatra kiterjedő biztosítások is sokkal olcsóbban kaphatók. Mivel ilyen veszélyességű terület nem kevés van Magyarországon, erre feltétlenül tekintettel kell lenni a díjak meghatározása során.

Az árvíz-kockázatok megfelelő felmérését csak hidrológiai, vízügyi mérnöki elemzé-sek, kutatások segítségével lehet elvégezni. Jelenleg hazánkban nagyon kis mértékben kü-lönböztetik meg árvíz-veszélyességi szempontból a biztosítási szerződéseket, lényegében csak bizonyos területeket zárnak ki a szerződéskötési lehetőségéből. Valójában az árvízi díjszükségletet szétterítik a teljes veszélyközösségre. Ez a porlasztás, áthárítás csupán ad-dig tehető meg, amíg nem jelenik meg egy olyan szereplő a piacon, amely az árvízzel nem veszélyeztetett területeken nem kínál olcsóbb biztosításokat.

A földrengéskockázatok megfelelő felmérése csak földrengéskutatók (geofizikusok) segítségével oldható meg. Ők rendelkeznek olyan földrengés-adatbázissal és geofizikai modellel, amellyel modellezhető a földrengések előfordulási helye, mélysége, magnitúdója és terjedése. Az adott intenzitású rengések által okozott károk meghatározásához külföldi biztosítási tapasztalatokat javasolt felhasználni, de tekintettel kell lenni arra, hogy az eltérő építési kultúra és szabályozás, esetleg egyéb tényezők miatt ezek magyarországi felhasz-nálhatósága csak korlátozott. A magyarországi települések földrengés veszélyeztetettsége eltérő, azonban az eltérések nem túl nagyok. Például az, hogy egy régi családi ház rendel-kezik-e betonkoszorúval sokkal fontosabb kockázati tényező, mint a ház területi elhelyez-kedése. Jelenleg nem tudok magyarországi díjmegkülönböztetésről földrengéskockázat miatt.

Hogyan határozzák meg a biztosítók a kockázati díjszükségletet? Módszerei nem nyil-vánosak. Azt lehet tudni, hogy egyes cégek aktuáriusi szoftvereket, mások saját fejlesztésű

programokat használnak. A leggyakoribb, hogy mind a kádarabszámra, mind a kárnyagra általánosított lineáris modellt (GLM) alkalmaznak. Ez persze azt feltételezi, hogy mind a területeknél, mind a koroknál már elvégezték a csoportosítást.

Nem életbiztosítási magyarországi területi becslésekről eddig a Márkus–Arató–Prokaj (2010) és Vitéz (2007) tanulmányai jelentek meg.

Biztosítási kockázatok területi becslésének szakirodalma

A területi kockázatok aktuáriusi-statisztikai vizsgálatát általában nagyon nehezítette a települések relative magas száma és így az ismeretlen paraméterek száma gyakran kezelhetetlennek bizonyult. Emiatt gyakran igazi statisztikai modellek alkalmazása nélkül, egyes jellemzők (pl. kárhányad, kárgyakoriság) segítségével történt a tarifabesorolás (az ilyen jellegű érdekesebb publikációk közül lásd például (Brubaker 1996, Christopherson–Werland 1996, Guven 2004, Wang–Zhang 2003)). A biztosítási kockázatok területi különbségeit tudományos igényességgel először Taylor vizsgálta (1989). Ausztráliai ingóságbiztosítások régiós szintű kárányaira és kárgyakoriságaira illesztett simításként kétdimenziós splineokat.

Boskov és Verrall publikálta az első olyan aktuáriusi cikket (1994), amely felhasználta a térbeli statisztika eredményeit. A szerződéseknek csak a földrajzi helyét ismerték és ennek hatását becsülték. A betegségtérképezéshez hasonlóan felismerték azt, hogy a területek nagy száma miatt a területenkénti önálló becslés nem adna megbízható becslést. A modell lényegében Besag, York és Mollié modellje (BYM) (1991) volt, tehát a kárszámokról Poisson-eloszlást feltételeztek, a kárszámintenzitásra egy lognormális modellt, ahol erőteljes a szomszédsági hatás. A modellben és a számításokban a károk számát és a szerződésszámokat (vagy az előzetes számításokban már meghatározott többi hatás szorzójával módosított szerződésszámokat) régióként (területenként) összesítették és ezekre az összesített értékekre számoltak.

Röviden bemutatjuk modelljüket. Az i -edik régió kárszámát jelöljük η_i -vel, feltételes eloszlásuk Poisson. Legyen λ_i a i -edik régió Poisson paramétere, amit a következő alakban keresünk:

$$\begin{aligned}\lambda_i &= e_i \cdot x_i, \\ \log(x_i) &= t_i + u_i + v_i.\end{aligned}\tag{1}$$

Az e_i a szerződésben töltött idő módosított értéke, t_i az i -edik régió ismert hatása, u_i a területi függőségi hatás, v_i a véletlen ingadozás. Ez utóbbiról feltételezzük, hogy azonos eloszlású 0 várható értékű független normálisak.

Ha ismerjük x_i -t, a károk számának eloszlása a következő módon adható meg:

$$P(\eta_i = y_i | x_i) = \frac{\exp(-e_i x_i) (e_i x_i)^{y_i}}{y_i!}\tag{2}$$

Térjünk rá a térbeli hatás változóinak eloszlására. A szerzők a BYM modellhez hasonlóan összefüggést tettek fel a szomszédos régiók között. Ehhez persze definiálni kell egy szomszédsági viszonyt a régiók között. Szomszédosnak lehet tekinteni a közös határral rendelkező régiókat, de településeknél gyakori például, hogy a valamilyen rögzített távol-

ságnál közelebbiek a szomszédok. Jelölje δ_i azon indexek halmazát, amely sorszámu régiók szomszédosak a j . régióval. Ekkor alkalmazható a Markov-mező-modell, amelyben az U_i -k a priori eloszlását, a σ^2 hiperparaméter mellett a következőnek választjuk:

$$[u_i | u_{j, j \neq i}, \sigma^2] \propto N\left(\frac{1}{|\delta_i|} \sum_{j \in \delta_i} u_j, \frac{\sigma^2}{|\delta_i|}\right).$$

ahol $[x | y]$ típusú jelölés folytonos esetben az $f_X(x | Y = y)$ feltételes sűrűségfüggvény, diszkrét esetben a $P(X = x | Y = y)$ helyett áll. $\propto$ -vel jelöljük az eloszlásban egyezőséget, vagy a konstans szorzótól eltekintve egyenlőt. U_i feltételes eloszlása tehát normális, a várható értéke a szomszédos U_j -k átlaga, szórása pedig a szomszédjai számával fordítottan arányos. Ekkor az U_i -k együttes eloszlása a következőképpen írható fel:

$$[u | \sigma^2] \propto \frac{1}{\sigma^2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^m \sum_{j \in \delta_i, j < i} (u_i - u_j)^2\right\}$$

A σ^2 hiperparaméter azt mutatja, hogy milyen mértékű hasonlatosságot tételezünk fel a szomszédos régiók között. Ha a σ^2 -t 0-nak választjuk, akkor a priori minden U_i azonos lesz, míg $\sigma^2 = \infty$ választással az egyes régiók hatása egymástól független. Persze általában nincs megbízható információ a régiók közti összefüggés mértékéről, gyakran σ^2 -t is valószínűségi változónak tekintjük.

Boskov és Verrall munkájának továbbfejlesztését jelentik azok a (Brouhns–Denuit–Masuy–Verrall 2002, Denuit–Lang 2004) publikációk, ahol megpróbálják egyszerre kezelni a különböző kockázati tényezőket. Denuit és Lang cikkének (2004) térbeli modellje megegyezik az előzővel, de itt szerződés szintű modellről van szó.

Az egyes szerződések kárszámát jelöljük Y_i -vel, feltételes eloszlásuk itt is Poisson. Amennyiben s kockázati tényező (például kor, nem, kocs típusa, lakóház magassága stb.) befolyásolja a kárdarabszámot, akkor az i . szerződő Poisson λ_i paraméterére és ennek logaritmusára feltételezték, hogy

$$\log(\lambda_i) = \log(e_i) + \beta_0 + \beta_{k_{1,i}}^1 + \dots + \beta_{k_{s,i}}^s + u_{r_i} + v_{r_i}.$$

Itt e_i a szerződésben töltött időt, $\beta_{k_{m,i}}^m$ az i . szerződőre jellemző m . kockázati csoport hatását jelöli. u_{r_i} és v_{r_i} az i . szerződéshez tartozó településnek, régióknak a hatását adja. Ezek eloszlása megegyezik Boskov és Verrall cikkében szereplővel.

Y_i -kről feltételezték, hogy feltételes függetlenek és Poisson eloszlásúak, így a Bayes-tétel segítségével felírható U_j -k a posteriori eloszlása:

$$[u_j | u_{k, k \neq j}, \sigma^2, \underline{\beta}, y] \propto \prod_{i=1, i \neq j}^n \exp\left\{-\exp(\log(e_i) + \beta_0 + \beta_{k_{1,i}}^1 + \dots + \beta_{k_{s,i}}^s + u_{r_i})\right\} \times \exp(y_i u_j) \times \exp\left(-\frac{1}{2\sigma^2} u' K u\right)$$

ahol K egy négyzetes mátrix (l a sorok száma megegyezik a régiók számával), amelynek elemei a főátlóban a megfelelő régió szomszédainak száma: $K_{r,r} = |\delta_r|$, a többi helyen $K_{r,s} = -1$, ha az r . és s . régió szomszédosak, egyébként 0. A σ^2 a már említett hiperparaméter, amely azt mutatja, hogy milyen mértékű hasonlatosságot tételezünk fel a szomszédos régiók között. σ^2 a priori eloszlását Lang és Denuit inverzgammaának választotta, elsősorban a számolások megkönnyítése végett. A hiperparaméter a posteriori eloszlását a

$$[\sigma^2 | \underline{u}, y] \propto [\underline{u} | \sigma^2][\sigma^2]$$

kifejezés adja meg, mely (a, b) paraméterű inverzgamma a priori eloszlás esetén inverzgamma lesz; $\left(a + \frac{l}{2}, b + \frac{1}{2}u'Ku\right)$ paraméterekkel.

Kárnagyságokra egy kissé meglepő módon normális eloszlást javasolnak. Ez azért meglepő, mert a biztosítási számításokban általában exponenciális, lognormális, Pareto, gamma káreloszlásokkal dolgoznak. A kockázati tényezők és a terület hatását a kárszámeloszlás paraméterének logaritmusára felírt modellel valósítják meg úgy, hogy ugyanezt a modellt a káreloszlás várható értékére írják fel.

A Dimakos–Frigessi (2002)-cikk teljes Bayesi megközelítést alkalmaz valamennyi kár- és kártényezőre, azonban itt a régiók száma mindössze 19. A káreloszlásnál a Denuit–Lang cikkel ellentétben feltételes gammaeloszlást illesztenek.

Mindkét publikációban a modell bonyolultsága miatt Markov-lánc Monte Carlo (MCMC)-eljárással becsülik meg a paramétereket.

A Márkus–Arató–Prokaj (2010) cikkben Boskov–Verrall (1)-(2) modelljéhez hasonlóan csak a területi hatást vizsgáltuk. A területi tényezőkön kívüli kockázati hatások a vizsgálat elején GLM-el lettek megbecsülve. Az egyszerűsítést az indokolta, hogy több mint 3000 településre becsültük meg a területi hatást.

Itt Clayton és Kaldor (1987) feltételes autoregressziós modelljének (CAR) egy modifikációját alkalmaztuk. Az i -edik település Poisson paramétere $\lambda \cdot e^{g_i} \cdot t_i$, ahol t_i a település szerződéseinek területi kockázatoktól különböző kockázataival módosított szerződésben töltött idejének összessége, λ a kárszám közös paramétere. A g_i területi együtthatók együttes kovarianciamátrixát

$$\Sigma = \tau^{-2} D(I - \rho A)^{-1} D$$

alakban kerestük, ahol I az egységmátrix, $A_{r,s} = 1$, ha az r . és s . régió (település) szomszédos és különbözőek, egyébként 0. D diagonális mátrix a főátlóban pozitív elemekkel. A ρ paramétert a $(0, \rho_{\max})$ intervallumon keressük, ahol $\rho_{\max}$ az A mátrix legnagyobb sajátértékének reciproka. A pozitivitást azért követeljük meg, hogy a kovariancia bármely két település között ne legyen negatív, a másik határ a kovarianciamátrix pozitív szemidefinitéséből adódik. Érdekeség, hogy amennyiben az összes többi paramétert rögzítjük, úgy a likelihoodfüggvény maximuma λ -ban könnyen adódott.

$$\hat{\lambda} = \frac{\left(\sum_{i=1}^N y_i\right)}{\sum_{i=1}^N e^{g_i} t_i}$$

A modell előnye, hogy a likelihoodfüggvényhez szükséges mátrixinvertálást ez esetben nem kell minden MCMC lépésben elvégezni, elég egyszer meghatározni az A mátrix sajátértékeit. Ez lehetővé teszi a számításokat még a több mint 3000x3000-es mátrixoknál is.

Gschlößl és Czado (2007) cikkében a kárszámok térbeli összefüggése szintén CAR modellel lett megközelítve, azonban a kovarianciamátrix egészen más alakú volt. A cikk újdonsága abban rejlett, hogy megvizsgálták a károk számának és a kár nagyságának összefüggését is. A káreloszlást gammának választották és magyarázó változónak bevonták a szerződés kárszámát is. A vizsgált német adatokon a várt eredményt kapták, tehát több kár esetén kisebb a várható kár nagysága.

Scheel és társai kutatásukban (2013) azt vizsgálták meg, hogy milyen összefüggés van kárszám, kárnagyság és meteorológiai körülmények között. A cél annak kiderítése volt, hogy az esetleges éghajlatváltozás milyen hatással lenne a biztosítókra.

Az (α, λ) paraméterű akadályozott Poisson eloszlás a következő:

$$P(N=0) = \alpha, \quad P(N=n) = (1-\alpha) \frac{\lambda^n e^{-\lambda}}{(1-e^{-\lambda})n!} = (1-\alpha) \frac{\lambda^n}{(e^\lambda - 1)n!}, \quad n = 1, 2, 3, \dots$$

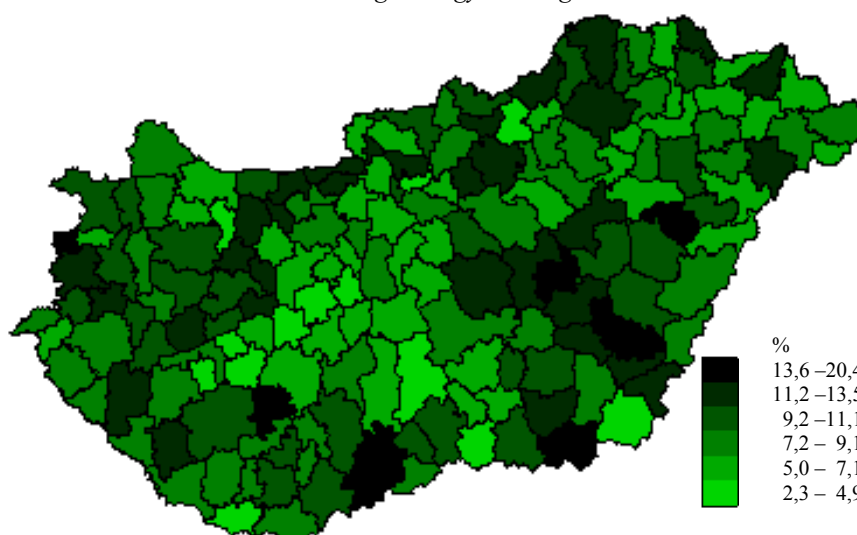
Norvégia régióinak napi kárszámaira ezt az akadályozott Poisson modellt írták fel úgy, hogy az első paraméter a terület aznapi és néhány korábbi napi meteorológiai adatától függött, a második paramétert pedig az állomány és a terület jellemzői határozták meg. Érdeklenség, hogy a területi összefüggést Ising-moddell határozták meg, azzal, hogy az Ising-modell egyik állapota azt jelzi, hogy a kockázati hatás érvényesül a területen, a másik állapot pedig a hatástalanságot jelzi.

Egy magyarországi biztosítási példa

A biztosítási kockázatok területi kockázataival kapcsolatos számításokra nézzünk meg egy magyarországi példát! A nem publikus adatbázis több mint 500 ezer szerződés egy speciális nem életkár típusának adatait tartalmazza, több mint 1 millió 800 ezer időegységnyi megfigyeléssel és több mint 150 ezer kárral. Nagyon fontos, hogy a szerződés- és káradatok szerződésenként álltak rendelkezésünkre. A szerződéseknél 3 kockázati tényezőt különböztettünk meg. Lakásbiztosításoknál ezek például lehetnének a lakóépület, a szerződő korcsoportja és a tetőzet fajtája. Gépjármű-biztosításnál pedig például a gépjármű típusa, a szerződő korcsoportja és az üzemanyag fajtája. 3171 település kockázatát szeretnénk megbecsülni. A településeket 168 kistérségbe soroltuk és 10 fajta település nagyságot különböztettünk meg.

1. ábra

Kistérségek kárgyakoriságai



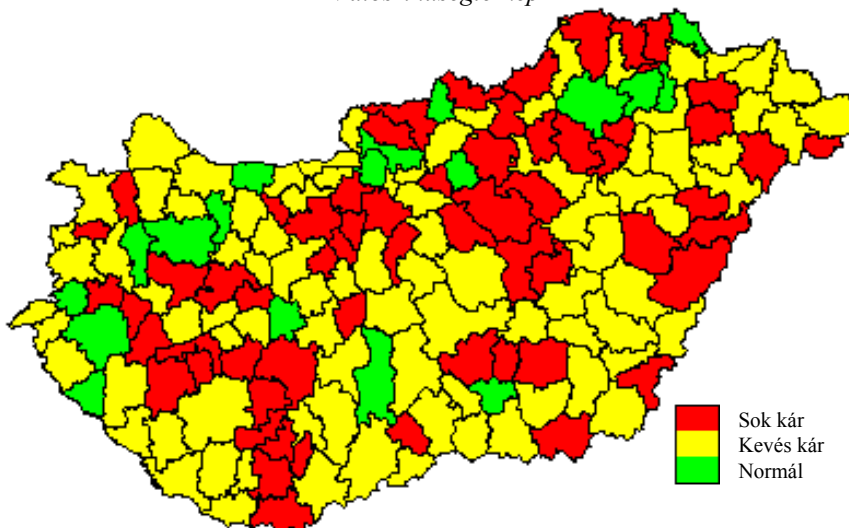
Forrás: saját szerkesztés.

A települések kárszáma nagyon különbözik egymástól. Amennyiben nagyobb területi egységeket, kistérségeket veszünk, akkor is láthatjuk az 1. ábrán, hogy a kárgyakoriság (károk száma osztva a kockázatban töltött idővel) 2,3%-tól 20,4%-ig változik.

A térbeli modellek ellenőrzésére az egyik első módszer, hogy ábrázoljuk azokat a területeket, ahol a kárszám vagy nem éri el a területre előrejelzett eloszlás 5%-os kvantilisét, vagy meghaladja a 95%-os kvantilist. A 2. ábrán egy ilyen térképet láthatunk (valószínűségi térkép vagy probability map), ha a kárszámról Poisson-eloszlást feltételezünk, de semmiféle területi hatás nélkül. Piros színnel jeleztük azokat a területeket, melyeknél a bekövetkezett kárszám meghaladja az eloszlás 95%-os kvantilisét, sárgával, ahol nem éri el az 5%-os kvantilist, és zölddel azokat a területeket, ahol a kárszám a középső 90%-os sávba esik. Az olyan térképeknél, ahol a piros és sárga területek részaránya jelentősen eltér az 5–5%-tól, különösen, ha jelentősen meghaladja azt, akkor azt mondhatjuk, hogy az alkalmazott modell nem felel meg a kockázatok területi eloszlásának. Az adott ábra pont ezt mutatja, mivel a kistérségek nagy része a „kis valószínűségű” részbe esik, így azt a következtetést vonhatjuk le, hogy a területi hatás igen jelentős.

2. ábra

Valószínűségi térkép



Forrás: saját szerkesztés.

Az egyes szerződések kárszámát továbbra is Y_i -vel jelöljük, feltételes eloszlásuk itt is Poisson, λ_i paraméterrel. Az i -edik szerződés kockázatban töltött tartama e_i . A következő modelleket illesztettük a kárszámadatokra.

1.) Általánosított lineáris modell (GLM), úgy, hogy a településjellemzők közül csak a településnagyságot vesszük figyelembe.

$$\log(\lambda_i) = \log(e_i) + \beta_0 + \beta_{k_{N,i}}^N + \beta_{k_{A,i}}^A + \beta_{k_{B,i}}^B + \beta_{k_{C,i}}^C.$$

Itt $k_{N,i}$, $k_{A,i}$, $k_{B,i}$, $k_{C,i}$ azt jelzik, hogy az i -edik szerződés települése nagyság szerint melyik csoportban van, illetve az A, B, C kockázat szerint a szerződésnek mi a jellemzője.

Felmerülhet a kérdés, hogy miért pont a településnagyságot vesszük figyelembe a településjellemzők közül. Ennek egyik oka, hogy meglepő módon még a természeti károknál is van hatása. Ennek magyarázata lehet egyrészt a településmérettől függő építkezési mód, másrészt kárbejelentési hajlandóság különbözősége is. Természetesen a konkrét esetekben mindig érdemes körülnézni az elérhető adatok között. Például a gépjármű-biztosításoknál lehet, hogy fontos a gépjárművek száma.

2.) Általánosított lineáris modell (GLM) úgy, hogy a faktorok közé bevesszük a kistérséget is

$$\log(\lambda_i) = \log(e_i) + \beta_0 + \beta_{k_{N,i}}^N + \beta_{k_{A,i}}^A + \beta_{k_{B,i}}^B + \beta_{k_{C,i}}^C + \alpha_{r(i)}.$$

Itt $r(i)$ azt mutatja, hogy az i -edik szerződés települése melyik kistérséghez tartozik.

3.) Paraméterek a priori eloszlása gamma.

$$\log(\lambda_i) = \log(e_i) + \beta_0 + \beta_{k_{N,i}}^N + \beta_{k_{A,i}}^A + \beta_{k_{B,i}}^B + \beta_{k_{C,i}}^C + \delta_{s(i)}.$$

Itt $s(i)$ az i -edik szerződés települését jelzi. Feltételezzük, hogy valamennyi β és δ paraméterre $\exp(\beta)$ -k és $\exp(\delta)$ -k a priori eloszlása független 1 várható értékű gamma eloszlás. Ebben az esetben könnyen számolható minden paraméter a posteriori várható értéke (a többi paramétert állandónak tekinthető), mivel

$$\frac{\text{eredeti gamma paraméter} + \text{csoport kárainak száma}}{\text{eredeti gamma paraméter} + \text{csoport kockázatba n töltött összeideje}}$$

törttel egyenlő.

4.) Paraméterek a priori eloszlása gamma úgy, hogy a faktorok közé bevesszük a kistérséget is.

$$\log(\lambda_i) = \log(e_i) + \beta_0 + \beta_{k_{N,i}}^N + \beta_{k_{A,i}}^A + \beta_{k_{B,i}}^B + \beta_{k_{C,i}}^C + \alpha_{r(i)} + \delta_{s(i)}.$$

5.) Kárgyakoriságok simítása. Algoritmusunk¹⁾ a következő lépésekből áll.

- Először a 168 kistérség kárgyakoriságát (károk száma osztva a kockázatban töltött idővel) határozzuk meg.
- A kistérségi kárgyakoriságokra megfelelő simaságú felületet illesztünk az R programcsomag `surf.gls` függvényével. Így valamennyi településre kapunk egy előzetes kárszámintenzitás becslést.
- A kockázatban töltött időket módosítjuk ezekkel az intenzitásokkal.
- Elvégezzük az 1. modellnek megfelelő GLM-számítást a módosított kockázati időkkel.
- Az eredeti kockázati időket módosítjuk a GLM-számításból kapott együtthatókkal.
- 4-szer megismételjük az a)-e) lépéseket.
- Az intenzitásértéket és a településnagysági szorzót összesítve a települési szorzókat pl. 5 csoportba soroljuk és ezt mint szempontot figyelembe véve újra elvégezzük a GLM-számítást.

A modell számításai nagyon gyorsak a szokásos területi számításokhoz képest.

Látható, hogy a területi összefüggéseket csak az 5. modellben használjuk, de a fejezetben csak az a célunk, hogy bemutassuk a területi számítások modelljeinek összehasonlítását.

1) Ez az algoritmus I.Drydennel, L. Márkussal és C. Taylорral közös munka.

Milyen módszerekkel lehet összehasonlítani a modelleket? Bayes-i modellek esetében igen gyakori, hogy a Spiegelhalter és társai (2002) DIC (deviance information criterium) kritériumot használják. Ez történt a Denuit–Lang (2004) és Gschlößl–Czado (2007) cikkben is. Amennyiben bayesi és nem bayes-i modellek is felmerülnek, úgy javasolható cross-validation (talán leggyakrabban keresztkiértékelésnek és kereszt-validálásnak szokták fordítani) alkalmazása.

Esetünkben az eljárás a következő.

- 1.) A több mint 3000 településből kiválasztunk 2000 települést. A véletlenszerű kiválasztás a kockázati időkkal súlyozva történik.
- 2.) A 2000 település szerződéseinek 70%-át választjuk ki visszatevés nélküli mintavétellel.
- 3.) A kiválasztott szerződések (gyakorló adatbázis) alapján elvégezzük a paraméterek becslését és a becsült paraméterekkel előrejelezzük a kimaradt szerződések (teszt adatbázis) kárszámának eloszlását. A tényleges kárszámok ismeretében kiértékeljük az eloszlások megfelelőségét.
- 4.) Az eljárást 100-szor megismétljük. Nagyon számítás- és időigényes modelleknél akár 2-3 ismétlés is igen hasznos lehet.

Az előrejelzéseket a következő mérőszámokkal értékeljük. Mivel a mostani vizsgálatunknál a legfontosabb területi hatás, ezért a számítások vagy településenként, vagy kistérségenként történnek meg és utána összesítünk az egész országra. z_j jelöli a j -edik település tényleges kárszámát a teszt adatbázisban és $\hat{z}_j$ ennek előrejelzését a gyakorló adatbázis alapján.

- 1.) Négyzetes eltérés:
$$\sum_j (z_j - \hat{z}_j)^2$$
- 2.) Módosított négyzetes eltérés:
$$\sum_j \frac{(z_j - \hat{z}_j)^2}{\hat{z}_j}$$
- 3.) Logaritmikus szkór:
$$-\sum_j \log \left(\frac{(\hat{z}_j)^{z_j} e^{-\hat{z}_j}}{z_j!} \right)$$
- 4.) Brier szkór:
$$-\sum_j \left(2 \frac{(\hat{z}_j)^{z_j} e^{-\hat{z}_j}}{z_j!} - \sum_{k=0}^{\infty} \left(\frac{(\hat{z}_j)^k e^{-\hat{z}_j}}{k!} \right)^2 \right)$$

A 3. és a 4. mérőszámnál kihasználtuk, hogy a cikkben Poisson-eloszlást feltételeztünk. Minél kisebbek ezek a mérőszámok, annál jobban találtuk el a települések kárszámának eloszlását.

Az 5 modell mérőszámai a következők lettek (a 100 cross-validation eljárás átlagában).

1. táblázat

A modellek mérőszámai

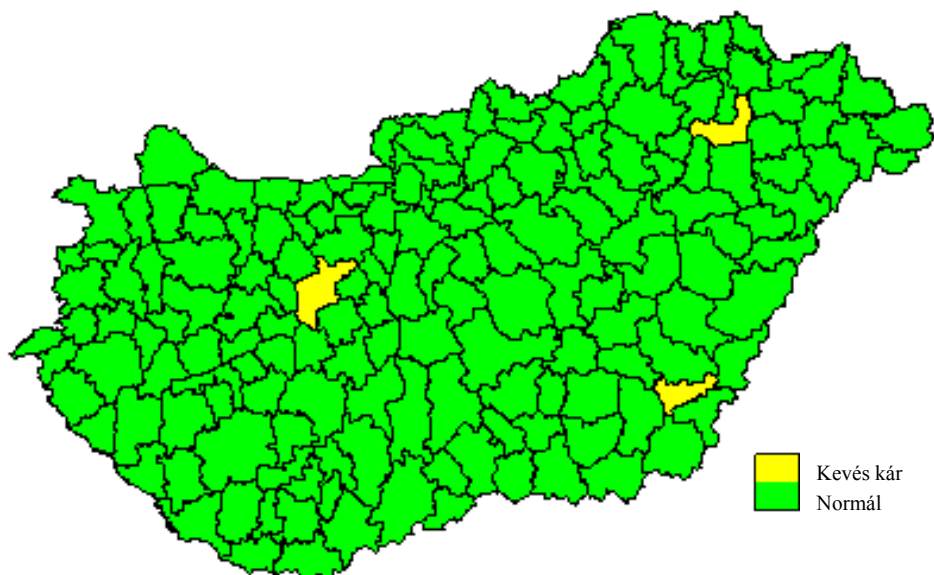
Modellek	Négyzetes eltérés	Módosított négyzetes eltérés	Logaritmikus szkór	Brier szkór
1. modell	727 884	9 677	9 013	32
2. modell	261 177	6 716	7 556	30
3. modell	150 195	5 560	7 051	18
4. modell	146 729	5 176	6 840	16
5. modell	261 715	6 834	7 640	29

Forrás: saját szerkesztés.

A táblázat azt mutatja, hogy a vizsgált modellek közül a valamennyi paraméterre apriori gamma eloszlást feltételező modell adta a legjobb eredményt.

Erre a modellre és az eredeti adatokra is ábrázoljuk a 3. ábrán a valószínűségtérképet. Az előző valószínűségtérképhez képest óriási a változás, a kiugró területek száma a töredékére csökkent. Látható, hogy ebben az esetben talán „túlzottan is jól” sikerült az illesztés, hiszen csak 3 kistérségnél van kiugróan alacsony érték, miközben körülbelül 17 nem zöld területet várnánk. A „túlzottan is jó” kifejezést úgy értjük, hogy az ábra azt sejteti, hogy túlparaméterezésről lehet szó, de ez nem feltétlenül van így. Figyelembe kell venni azt, hogy a kvantilisok és a tényleges káradatok összehasonlítását ugyanazokon az adatokon végeztük el, amelyek alapján a paramétereket becsültük. Ilyen esetben általában túl pontos illeszkedést kapunk. A keresztkiértékelésnél kapott számok alapján mindenképpen ezt a modellt érdemes preferálni.

3. ábra

Valószínűségtérkép gamma a priori eloszlással

Forrás: saját szerkesztés.

Összefoglalás

A tanulmányban bemutatunk több olyan területi statisztikai modellt, amelyeket biztosítási alkalmazásokban használnak fel. Ezek nagyobb része kárszámokra vonatkozott, de voltak kárnagyságokra illesztettek is. Egy konkrét magyarországi biztosítási adatbázis kapcsán bemutatunk néhány területi függést nem feltételező struktúrát is. Itt hangsúlyoztuk az egyedi adatok felhasználásának fontosságát, hiszen ezáltal alkalmazhatunk például keresztkiértékeléseket is. A modellek összehasonlítását ezzel a keresztkiértékeléssel és előrejelzési mérőszámok (szkórok) felhasználásával végeztük el. A bemutatott példában így találtunk is egy a többinél egyértelműen jobb modellt.

Annak ellenére, hogy a bemutatott módszertan és példa biztosítási, a tanulmányban ismertetett eszközök felhasználhatók minden olyan területen is, ahol területi függéseket feltételezhetünk és egyedi szintű adatbázis áll rendelkezésre.

IRODALOM

- Arató, N. M.–Dryden, I. L.–Taylor, C. C. (2006): Hierarchical Bayesian modelling of spatial age-dependent mortality *Computational Statistics & Data Analysis* 51 (2): 1347–1363.
- Márkus, L.–Arató, M.–Prokaj, V. (2010): Hierarchical Bayesian Modelling of Geographic Dependence of Risk in Household Insurance In: Skiadas, C. (Ed.): *Advances in Data Analysis: Theory and Applications to Reliability and Inference Data Mining, Lifetime Data and Neural Networks* pp. 219–227., Birkhäuser, Basel.
- Bálint, L. (2010): A területi halandósági különbségek alakulása Magyarországon 1980–2006, (Központi Statisztikai Hivatal Népeségstudományi Kutatóintézetének Kutatási Jelentései; 90.), KSH Népeségstudományi Kutató Intézet, Budapest.
- Besag, J.–York, J.–Mollié, A. (1991): Bayesian image restoration, with two applications in spatial statistics *Annals of the Institute of Statistical Mathematics* 43 (1): 1–20.
- Boskov, M.–Verrall, R. J. (1994): Premium rating by geographic area using spatial models *ASTIN Bulletin* 24 (1): 131–143.
- Brouhns, N.–Denuit, M.–Masuy, B.–Verrall, R. (2002): Ratemaking by Geographical Area: A Case Study Using The Boskov and Verrall Model *Université Catholique De Louvain, Institut De Statistique, Discussion Paper* 0202, Louvain.
- Brubaker, R. E. (1996): Geographic Rating of Individual Risk Transfer Costs without Territorial Boundaries *Casualty Actuarial Society Forum*, Winter 1996. pp. 97–127.
- Christopherson, S.–Werland, D. L. (1996): Using a Geographic Information System to Identify Territory Boundaries *Casualty Actuarial Society Forum*, Winter 1996. pp. 191–211.
- Clayton, D.–Kaldor, J. (1987): Empirical Bayes estimates of age-standardized relative risks for use in disease mapping *Biometrics* 43 (3): 671–681.
- Denuit, M.–Lang, S. (2004): Non-life rate-making with Bayesian GAMs *Insurance: Mathematics and Economics* 35 (3): 627–647.
- Dimakos, X. K.–Frigessi, D. R. A. (2002): Bayesian Premium Rating with Latent Structure *Scandinavian Actuarial Journal* 2002 (3): 162–184.
- Gschlöbl, S.–Czado, C. (2007): Spatial modelling of claim frequency and claim size in non-life insurance *Scandinavian Actuarial Journal* 2007 (3): 202–225.
- Serhat, G. (2004): Multivariate Spatial Analysis of the Territory Rating Variable, in 2004 Discussion Paper Program - Applying and Evaluating Generalized Linear Models Including Research Papers on the Valuation of P&C Insurance Companies *Casualty Actuarial Society* pp. 245–260.
- Major, J. A.–Yakov, L. (2001): Actuarial Applications of Multifractal Modeling Part I. Introduction and Spatial Applications <http://www.casact.org/pubs/forum/01wforum/01wf321.pdf> (letöltve: 2015. augusztus).

- Scheel, I.–Ferkingstad, E.–Frigessi, A.–Haug, O.–Hinnerichsen, M.–Meze-Hausken, E. (2013): A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62 (1): 85–100.
- Spiegelhalter, D.–Best, N.–Carlin, B.–van der Linde, A. (2002): Bayesian measures of model complexity and fit *Journal of the Royal Statistical Society: Series B* 64 (4): 583–640.
- Taylor, G. C. (1989): Use of spline functions for premium rating by geographic area *ASTIN Bulletin* 19 (1): 89–122.
- Vitéz, I. (2007): Location as risk factor Spatial analysis of an insurance data-set In: Skiadas, C. H.: *Recent Advances in Stochastic Modeling and Data Analysis* pp. 130–136., World Scientific Publishing, Singapore.
- Wang, H. H.–Zhang, H. (2003): On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach *The Journal of Risk and Insurance* 70 (1): 111–124.

Kulcsszavak: területi statisztika, biztosítási kárszám, betegségtérképezés, BYM modell, Markov véletlen mező.

Resume

Regional differentiation of insurance fees looks back upon a relatively long past, nevertheless, systematic actuary examination of regional risks has relatively short history. First publications appeared at the 1980s of last century inspired by sickness mapping models.

The present paper introduces several such regional statistical methods, that are good for analysing insurance risks. Actuary work is presented through an insurance example in Hungary, which also takes into consideration regional effects.