

Szakdolgozat

Szabadsfalvi Dániel

2022

Eötvös Loránd Tudomány Egyetem
Természettudományi Kar

Budapesti Corvinus Egyetem



Vigasráták modellezése

Elemzés a korrelációs kapcsolat és a bimodális
tulajdonság segítségével

Készítette: Szabadjfalvi Dániel

Biztosítási és pénzügyi matematika MSc szak

Kvantitatív pénzügyek specializáció

2022

Szakszemináriumvezető: Dr. Havran Dániel

NYILATKOZAT

Név: Szabadfalvi Dániel

ELTE Természettudományi Kar, szak: Biztosítási és pénzügyi matematika MSc szak

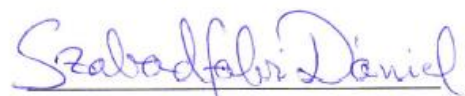
NEPTUN azonosító: FHOSZ6

Diplomamunka címe:

Vigasráták modellezése: Elemzés a korrelációs kapcsolat és a bimodális tulajdonság segítségével

A **diplomamunka** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2022. május 27.



a hallgató aláírása

Tartalomjegyzék

1. Bevezetés	1
2. A vigaszára modellezésének fő kérdései	4
2.1. Szakirodalmi előzmények	4
2.2. Korrelációs (wrong-way risk) modellek	8
2.3. A vigaszára eloszlásának vizsgálata regressziós módszerekkel.....	17
2.4. Két módszertan összehasonlítása	26
3. Adatbázisok bemutatása	28
3.1. A Bondora adatbázis bemutatása	28
3.2. Az Eurostat adatbázis bemutatása	30
3.3. Az összekapcsolt adatbázis bemutatása	32
3.4. Bővített adatbázis bemutatása (bimodalitás vizsgálata).....	33
4. Empirikus kutatás és modellezés	36
4.1. Adatbázisok elemzése	36
4.2. Wrong-way risk modellezés (korreláció vizsgálata).....	41
4.3. Regressziós modellezés (bimodalitás vizsgálata).....	55
5. Összefoglalás, konklúzió	64
Hivatkozásjegyzék	67
Adatforrások jegyzéke	71
Függelék	72

Ábrák jegyzéke

1. ábra: A negyedéves empirikus PD változása időben az egyes országok esetén	36
2. ábra: A negyedéves átlagos RR változása időben az egyes országok esetén.....	37
3. ábra: A vigaszáta tapasztalati sűrűségfüggvénye az egyes esetekben (országokén)	39
4. ábra: A vigaszáta tapasztalati sűrűségfüggvénye országonként (együttesen ábrázolva)	40
5. ábra: A vigaszáta hosszútávú eloszlásfüggvényének inverzfüggvénye.....	42
6. ábra: Az omega paraméter konvergenciája és eloszlása (az észt adatsor alapján).....	48
7. ábra: A rho1 paraméter konvergenciája és eloszlása (az észt adatsor alapján).....	48
8. ábra: A rho2 paraméter konvergenciája és eloszlása (az észt adatsor alapján).....	49
9. ábra: Az omega paraméter konvergenciája és eloszlása (a finn adatsor alapján)	50
10. ábra: A rho1 paraméter konvergenciája és eloszlása (a finn adatsor alapján).....	50
11. ábra: A rho2 paraméter konvergenciája és eloszlása (a finn adatsor alapján).....	51
12. ábra: Az omega paraméter konvergenciája és eloszlása (a spanyol adatsor alapján)	51
13. ábra: A rho1 paraméter konvergenciája és eloszlása (a spanyol adatsor alapján).....	52
14. ábra: A rho2 paraméter konvergenciája és eloszlása (a spanyol adatsor alapján).....	52
15. ábra: Vigaszáta eloszlásának becslése lineáris regresszióval (teljes adatbázis alapján)...	56
16. ábra: Vigaszáta eloszlásának becslése lineáris regresszióval (észt adatok alapján)	56
17. ábra: Vigaszáta eloszlásának becslése lineáris regresszióval (finn adatok alapján).....	57
18. ábra: Vigaszáta eloszlásának becslése lineáris regresszióval (spanyol adatok alapján) ...	57
19. ábra: Vigaszáta eloszlásának becslése GLM regresszióval (teljes adatbázis alapján).....	58
20. ábra: Vigaszáta eloszlásának becslése GLM regresszióval (észt adtok alapján)	59
21. ábra: Vigaszáta eloszlásának becslése GLM regresszióval (finn adtok alapján).....	59
22. ábra: Vigaszáta eloszlásának becslése GLM regresszióval (spanyol adtok alapján).....	60
23. ábra: Vigaszáta eloszlásának becslése béta regresszióval (teljes adatbázis alapján)	61
24. ábra: Vigaszáta eloszlásának becslése béta regresszióval (észt adatok alapján)	61
25. ábra: Vigaszáta eloszlásának becslése béta regresszióval (finn adatok alapján)	62
26. ábra: Vigaszáta eloszlásának becslése béta regresszióval (spanyol adatok alapján)	62

Táblázatok jegyzéke

1. táblázat: Wrong-way risk modellek összehasonlítása	16
2. táblázat: Az RR-t és a LGD-t becslő regressziós modellek összehasonlítása	25
3. táblázat: A modellezési módszertan összehasonlítása.....	26
4. táblázat: Főkomponenssúlyok táblázata (észt adatok)	44
5. táblázat: Főkomponenssúlyok táblázata (finn adatok)	44
6. táblázat: Főkomponenssúlyok táblázata (spanyol adatok)	45
7. táblázat: ARMA modellspecifikáció az egyes változók esetén (országoként).....	45
8. táblázat: Nemteljesítés valószínűségének (arányának) hosszútávú értéke az egyes esetekben (országok).....	46
9. táblázat: Összegző statisztikai táblázat wrong-way risk modellhez (Észtország).....	72
10. táblázat: Összegző statisztikai táblázat táblázat wrong-way risk modellhez (Finnország) 72	
11. táblázat: Összegző statisztikai táblázat táblázat wrong-way risk modellhez (Spanyolország).....	73
12. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Teljes adatbázis) 73	
13. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Észtország).....	74
14. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Finnország)	74
15. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Spanyolország) .	74

1. Bevezetés

A Bázeli II. szabályozás (Basel Committee on Banking Supervision, 2005) hatására a hitelkockázatok (*credit risk*) tőkekövetelményeinek számításánál meghonosodott a belső értékalapú (*IRB – internal rating-based*) megközelítés. A hitelkockázatok témakörének legfontosabb fogalma a nemteljesítési (*default*) esemény, mely legáltalánosabb definíciója szerint akkor következik be, ha a pénzügyi szerződés (például hitelszerződés, értékpapír, derivatív ügylet) partnere nem teljesít szerződés szerűen. A hitelszerződések esetén, a bázeli definíciók alapján mindez akkor következik be, ha a partner az esedékességet (az a nap, amikor a szerződés szerinti követelést ki kell egyenlítenie) követő 90 napban nem fizet.

Az IRB megközelítés lehetőséget biztosít a megszabott feltételeket teljesítő bankok számára, hogy a saját kockázati modelljeiket és értékeléseiket használják a szabályozói tőke kiszámításához. E megközelítés szerint a bankoknak meg kell becsülniük a következő kockázati összetevőket:

- a) **Nemteljesítés valószínűsége (*PD: probability of default*):** annak a valószínűsége, hogy adott időtávon (pl. egy év) belül egy adott partner esetében nemteljesítési (*default*) esemény következik be, azaz a partner nem tud a szerződésben meghatározott módon a fizetési kötelezettségeinek eleget tenni.
- b) **Nemteljesítés kitettsége (*EAD: exposure of default*):** a nemteljesítési esemény időpontjában (várhatóan) az adott partnerrel szemben fennálló teljes követelés mértéke (becsült érték).
- c) **Nemteljesítés miatti veszteség (*LGD: loss given default*):** a nemteljesítési eseményt követően befolyó pénzmennyiség (behajtási folyamat: *recovery process*) után fennmaradó veszteség százalékban kifejezett mértéke.
- d) **Tényleges szerződéses lejárat (*M: maturity*):** abban az esetben, ha a lejáratot kockázati komponensként kezelik, akkor a bankoknak meg kell adniuk a felügyeletnek a kitettségeik tényleges szerződéses lejáratát.

Mivel a Bázeli II. tőkekövetelmény-számítás erősen függ az LGD-től, a pénzügyintézetek az elmúlt években nagyobb hangsúlyt fektettek az LGD viselkedésének megfigyelésére és ezek alapján a kockázati tényező modellezésére. Ellentétbe a PD becsülésével, amelyre már jól bevált eszközeink vannak, az LGD nem annyira kézzel fogható, így még napjainkban is kutatás tárgyát

képezi. Szakdolgozatomban ezért az LGD, illetve az RR becslésével és modellezésével kívánok foglalkozni.

Az LGD-vel foglalkozó modellek fejlesztése segítséget nyújt a pénzüintézetek számára, hogy pontosabban fel tudják mérni a kockázataikat és szabályozói tőkekövetelményüket, valamint az adósságkezelésük javításában is fontos szerepet játszik.

Az LGD azon pénzüösszegek arányát jelenti, amelyet a pénzüintézetek nem tudnak beszedni a behajtási időszak alatt a hitelfeltevő nemteljesítése okán (*default-os*). Ezzel szemben a vigaszráta (*RR: recovery rate*) a pénzüintézetek által sikeresen behajtott pénzüösszegek, mínusz a beszedési időszak alatt keletkezett adminisztrációs díjak aránya a nemteljesítéskor fennálló teljes kitétséghöz (tartozáshoz) képest. A *recovery rate*-nek, nincs egyértelmű magyar megfelelője. A publikációkban gyakran használják a „visszaszerzési arány” kifejezést, ugyanakkor én a továbbiakban a „vigaszráta” fogalmat fogom használni a szakdolgozatomban. Az (1) és (2) egyenletek formális definíciókat adnak az RR-nek és az LGD-nek:

- Tegyük fel, hogy i hitelfeltevő vagy kötvényes már nemteljesítő (*default-os a hitele*), legyen EAD_i a nemteljesítéskori kitétsége az i -nek.
- Legyen A_i az i hitelfeltevő esetén felmerülő adminisztrációs költségek összege (például levelek, telefonhívások, ügyvédi díjak és egyéb jogi munkák költségei).
- Legyen R_i az i hitelfeltevő esetén behajtott összeg.

Ekkor

$$(1) \quad RR = \frac{R_i - A_i}{EAD_i} = \frac{\sum \text{Behajtott összegek} - \sum \text{Adminisztrációs költségek}}{\text{Nemteljesítéskor fennálló teljes tartozás}}$$

$$(2) \quad LGD = 1 - RR = 1 - \frac{R_i - A_i}{EAD_i}$$

A vigaszráta értéke általánosságban a $[0; 1]$ intervallumban található és jellemzően a 0 és 1 határpontokban koncentrálódik. Lehetséges, hogy RR negatív értéket vegyen föl, ekkor a behajtott pénzümennyiségek (*recoveries*) összege kevesebb, mint az adminisztrációs költségek értéke ($R_i < A_i$), de lehet nagyobb is, mint 1, ha a behajtott pénzümennyiség összértéke meghaladja a nemteljesítéskor fennálló tartozás és az adminisztrációs költségek összegét ($R_i > EAD_i + A_i$). Ezzel szemben modellezéskor az elmezők azzal a feltételezéssel élnek, hogy az RR (illetve az LGD) értéke általánosságban a $[0; 1]$ intervallumon mozog, ezért ezen az intervallumon érdemes becsülni is (csak ezen az intervallumon vehet föl értékeket).

Az RR becslésének egyik fő nehézsége a sűrűségfüggvényének bimodális tulajdonságában rejlik. Egy sűrűségfüggvény bimodális vagy multimodális, ha legalább két lokális maximumhelye van. Esetünkben a két maximumhely jellemzően az intervallum két szélén található, mert az RR értékei 0 és 1 körül koncentrálódnak. Az RR eloszlása mögött az az

empíria rejlik, hogy az emberek általában vagy teljes mértékben vissza tudják fizetni az adósságukat vagy közel semmit sem tudnak kifizetni belőle *default* esetén.

A szakdolgozatomban a vígaszrátához és nemteljesítés miatti veszteséghez kapcsolódó két fő (üzleti) modellezési kihívást – rosszirány kockázat és bimodális eloszlás – fogom ismertetni. Ehhez áttekintem a szakirodalom által ismert módszereket és adatokon keresztül kívánom bemutatni a működésüket. Ezáltal szemléltetni tudom a módszerek alkalmazhatóságának korlátait és használatukkal kapcsolatos aggályokat. Dolgozatomban az alkalmazott módszertanra fókuszálok és a fő kérdésem az, hogy milyen feltevések és körülmények között alkalmazhatók gyakorlatban a szakirodalom által ajánlott újnak tekinthető modellek és módszerek.

2. A vigaszráta modellezésének fő kérdései

Ebben a fejezetben áttekintem, hogy milyen módszertani fejlődés előzte meg a legújabb vigaszrátaival és nemteljesítés miatti veszteséggel foglalkozó modellek kialakulását. Ezt követően részletesen bemutatom a két modellezési területhez kapcsolódó gyakorlatorientált módszereket az utóbbi húsz év főbb publikációinak ismertetésével. Ezután e két módszertant összehasonlítom az általam kialakított szempontrendszer szerint.

2.1. Szakirodalmi előzmények

A legkorábbi hitelkockázatokkal kapcsolatos modell megalkotása az amerikai közgazdász, Robert Merton munkásságához köthető, aki a vállalat strukturális hitelkockázatát úgy modellezte, mintha a társaság saját tőkéje a vállalat eszközeire képzett vételi opció lenne. Ez kezdetben kifejezetten jó mérőszámnak bizonyult. Külön érdekességként emelem ki, hogy ezt a modellt bővítette ki és fejlesztette tovább Fischer Black és Myron Scholes, amiért később Nobel-díjat kaptak (Black & Scholes, 1973).

A Merton modell mögött meghúzódó gondolatmenet viszonylag egyszerű: nemteljesítésről akkor beszélhetünk, ha a vállalat eszközeinek piaci értéke alacsonyabb szintre süllyed, mint a vállalat kötelezettségeinek mértéke (Merton, 1974). A hitelezők a hitelviszonyt megtestesítő értékpapír (kötvény) lejártakor az adósság névértéke és a vállalat eszközeinek piaci értéke közül a kisebb mennyiséget kapják meg. Feltételezve, hogy a vállalat adósságát teljes egészében zérókupon kötvények alkotják.

Ha a vállalat lejáratkori értéke nagyobb, mint a kötvény névértéke, akkor a kötvénytulajdonosok visszakapják a kötvény névértékét. Azonban, ha a vállalat értéke kisebb, mint a kötvény névértéke, akkor a részvényesek nem kapnak semmit, a kötvénytulajdonosok pedig megkapják a vállalat piaci értékét. Így a kötvénytulajdonos számára a lejáratkor esedékes kifizetés megegyezik a kötvény névértékével, mínusz a cég értékére vonatkozó eladási opcióval, melynek a kötési ára megegyezik a kötvény névértékével, a lejárat, pedig a kötvény lejáratával.

Ebben a modellben az összes releváns hitelkockázati elem – az üzleti kockázat (az eszközök volatilitása) és a pénzügyi kockázat (a tőkeáttétel) – a vállalat strukturális

jellemzőinek függvényeként jelenik meg. Ebben az esetben a vigaszráta egy endogén, belső változó, mivel a hitelezők kifizetésfüggvénye a nemteljesítő (*default-os*) vállalat maradványértékétől függ. Merton elméletében a PD és az RR fordítottan arányosak. Például, ha a vállalat értéke nő, akkor a PD értéke csökken, eközben pedig az RR nemteljesítéskori várható értéke nő (*ceteris paribus*).

Ha kategorizálni szeretnénk a Merton modellt, akkor a strukturális modellek közé lehetne őt elhelyezni. A strukturális modellek lényege, hogy egy közgazdaságtani gondolatot állítanak gazdasági entitások, általában vállalatok hiteleinek bedőlése mögé magyarázatként. Ebbe a csoportba tartozik például a KMV, a Black-Cox és a Leland modell is.

A strukturális hitelkockázati modellek következő generációja a Merton által kidolgozott elméletre épített, azonban elvetették azt az irreális feltevést, hogy *default* esemény csak lejáratkor következhet be. Ehelyett feltételezték, hogy a nemteljesítési esemény a kibocsátás pillanatától, a lejáratig bármikor bekövetkezhet. A másik fontos különbség, hogy a nemteljesítési esemény pontosan akkor következik be, amikor a vállalat eszközeinek a piaci értéke egy előzetesen meghatározott (ismert) alacsony küszöbszint alá esik, szemben a korábbi feltételezéssel, hogy a *default* esemény akkor következik be, ha a vállalat eszközeinek piaci értéke a vállalat kötelezettségeinek összértéke alá esik. Az ilyen modellek közé tartoznak a Kim, Ramaswamy és Sundaresan (I. J. Kim és mtsai., 1993), Hull és White (Hull & White, 1995), valamint Longstaff és Schwartz (Longstaff & Schwartz, 1995) által megalkotott modellek.

A második generációs strukturális modellek esetén a vigaszráta értéke nemteljesítéskor egy külső, exogén változó, ami független a vállalat eszközértékétől. Általában a fennálló tartozás egy rögzített százalékaként határozzák meg, így a vigaszráta értéke független a nemteljesítés valószínűségétől.

A strukturális modellek hiányosságainak leküzdése érdekében létrejöttek az intenzitás alapú (redukált) modellek. Ezek közé tartozik a Litterman és Iben (Litterman & Iben, 1991), Madan és Unal (Madan & Unal, 1998), Jarrow és Turnbull (Jarrow & Turnbull, 1995), Jarrow, Lando és Turnbull (Jarrow és mtsai., 1997), Lando (Lando, 1998), Duffie és Singleton (Duffie & Singleton, 1999) és Duffie (Duffie, 1998) modelljei. A strukturális modellekkal ellentétben a redukált modellek a nemteljesítési eseményt nem a vállalat értékének alakulása alapján állapítják meg.

A redukált modellek külön explicit feltevéseket vezettek be mind a PD, mind az RR dinamikájára vonatkozóan. Ezeket a változókat a vállalat strukturális jellemzőitől (eszközeinek volatilitásától és tőkeáttételétől) függetlenül modellezték. Általánosan elmondható a redukált

modellekről, hogy a vigaszráta értékét exogén módon határozzák meg, amely független a nemteljesítés valószínűségétől.

Minden időpillanatban pozitív valószínűsége van annak, hogy a vállalat nem tudja teljesíteni a kötelezettségeit (*default* esemény). Így mind a nemteljesítés valószínűsége, mind az vigaszráta értéke sztochasztikusan változhat az idő függvényében (PD és RR egymástól független sztochasztikus valószínűségi változók).

E sztochasztikus folyamatok határozzák meg a hitelkockázat árát. Bár közvetlenül ezen értékek nem kapcsolódnak a vállalat eszközeinek értékéhez, de feltételezték, hogy van köztük valamilyen közvetett kapcsolat. Ezt az alternatív megközelítést nevezte el Duffie és Singleton (Duffie & Singleton, 1999) redukált formájú (intenzitási) modellnek.

A redukált modellek alapvetően különböznek a nemteljesítési események kiszámíthatóságával kapcsolatban a strukturális modellektől, mert az intenzitási modellek már képesek kezelni a hirtelen (meglepetésszerűen) bekövetkezett nemteljesítési eseményeket is. Egy tipikus redukált modell esetén feltételezzük, hogy a nemteljesítési esemény valószínűségét – amely minden időpillanatban pozitív értéket vesz föl – egy exogén valószínűségi változó határozza meg.

A nemteljesítési esemény akkor következik be, ha az intenzitási változó értéke egy előre meghatározott érték alá esik. A nemteljesítési eseményt egy előre nem látható Poisson eseményként kezelték. Ebben az esetben az a lényeg, hogy a ma rendelkezésre álló információk alapján, nem tudjuk előrejelezni, hogy mikor fog az intenzitási változó – innen a modell neve – értékében bekövetkezni ez a változás (ugrás).

A kilencvenes évek második felében a bankok és tanácsadó cégek elkezdtek a saját hitelkockázati modelljeik kidolgozását, amelyek célja az volt, hogy egy előre meghatározott megbízhatósági szint (konfidencia intervallum) mellett föl tudják mérni a potenciális veszteségeket, amelyet egy hitelkockázatnak kitett portfólió egy meghatározott időhorizonton (általában egy éven) belül szenvedhet el. A kockázatosított érték (*VaR: Value at Risk*) modellek közé tartozik a J. P. Morgan CreditMetrics (Gupton és mtsai., 1997), a Credit Suisse Financial Products CreditRisk+ (Credit Suisse Financial Products, 1997) és a McKinsey CreditPortfolioView (Wilson, 1998) modelljei.

A Credit VaR modellek nagyrészt redukált modelleknek tekinthetők, ahol a vigaszrátát jellemzően egy exogén állandó paraméterként vagy a nemteljesítési valószínűségtől független sztochasztikus változóként adták meg. Például a CreditMetrics, a CreditPortfolioView és a CreditPortfolioManager a vigaszrátát a nemteljesítési esemény valószínűségtől független sztochasztikus változóként kezeli – általában béta eloszláson keresztül modellezve. Mások,

például a CreditRisk+ állandó paraméterként kezelik az RR-t, amelyet minden egyes hitelkockázati kitétség esetén egy bemeneti értéként kell megadni.

A 2000-es évek második felében azonban átkerült a hangsúly a vigaszráták modellezésénél az elméleti kérdésekről a konkrét hiteleszközökhöz kapcsolódó RR és LGD becslésére. Ez a megközelítés sokkal specifikusabb és gyakorlatiasabb volt, így a modellezés valóságosabbá vált. Itt két nagyon fontos empirikus eredményt emelnék ki, amelyek valós adatok alapján mutatták be az RR és a PD közötti (a két folyamat között fennálló) kapcsolatot.

Altman, Resti és Sironi (Altman és mtsai., 2004) empirikus elemzése kimutatta, hogy a nemteljesítési (*default*) esemény valószínűsége és vigaszráta (*recovery rate*) negatívan korrelált, vagyis a két folyamat nem független egymástól, mint ahogy azt több korábbi elméleti modell állította. A második fontos eredményre a 2007-ben kitört válság elemzése során jöttek rá a kutatók (Chava és mtsai., 2011). Eszerint miközben a nemteljesítési esemény valószínűségei növekedtek, közben a vigaszráták szignifikánsan csökkentek (ezt nevezik rosszirány kockázatnak vagy *wrong-way risk*-nek).

Ezek az eredmények jól mutatják, hogy milyen fontos és összefüggő ez a két dimenzió a hitelkockázatok területén. A nemteljesítési valószínűség (PD) modellezése már több évtizede a hitelbírálati rendszerek eleme. A vigaszráták (RR) ugyanakkor az elmúlt másfél-két évtizedben kerültek a szakmai érdeklődés fókuszába. A vigaszráták fontosságát felismerve a kutatók és gyakorlati szakemberek számos publikációban vizsgálták a behajtási folyamatok (*recoveries*) viselkedését nemteljesítés esetén hitelkockázati kvantitatív modellek segítségével.

A vigaszrátával (RR) és a nemteljesítés miatti veszteségekkel (LGD) foglalkozó modern módszereket két nagy csoportba lehet osztani. Az első csoportba a korrelációs (*wrong-way risk*) modellek tartoznak. E modellek a PD és az RR lineáris kapcsolatából (lineáris korreláció) indulnak ki és pénzügy matematikai módszerek segítségével becsülik a nemteljesítési valószínűség és a vigaszráta, illetve nemteljesítés miatti veszteség között fellépő korreláció mértékét. A második csoportba olyan regressziós modellek (módszerek) tartoznak, melyek a vigaszráta és a nemteljesítés miatti veszteség sűrűségfüggvényének bimodális tulajdonságát próbálják megragadni (mindamellet, hogy pontos becslést és előrejelzést adnak az RR és LGD értékekre). Ezáltal lehetővé téve a vigaszráták és nemteljesítés miatt fellépő várható veszteség eloszlásának becslését.

A következő két alfejezetben e két módszertani csoportot fogom bemutatni. Fölvázolom a két módszer előnyeit, illetve nehézségeit. Mindezt a vonatkozó releváns szakirodalom néhány fontosabb publikációjának feldolgozásával teszem meg. Ezt követően pedig összehasonlítom a két modellezési módszertant a 2.4. alfejezetben.

2.2. Korrelációs (*wrong-way risk*) modellek

A korábbi feltételezésekkel ellentétben, a kitettség értékének alakulása (EAD) és a partner csőd kockázata (PD) nem független. Belátták, hogy a csőd valószínűség és a vigaszára mértéke negatívan korrelál, azaz a kapcsolat nem csak a PD és az EAD alakulás között figyelhető meg, hanem a PD és az RR közt is (Altman és mtsai., 2004).

A bázeli szabályozás „speciális *wrong-way risk*-ként” definiálja az általános csőd valószínűség és a kitettség nagysága közötti pozitív korrelációt. „Általános *wrong-way risk*-ként” pedig, ha a csőd valószínűség pozitívan korrelál a mögöttes piaci faktorokkal. Ebből kifolyólag az általunk vizsgált korrelációs modelleket is lehet *wrong-way risk* modelleknek nevezni.

A nemteljesítési esemény valószínűsége (PD) és a vigaszára (RR) közötti kapcsolatot kezdetben egy-, illetve két-faktoros modellek segítségével vizsgálták a kutatók, ahol az egyik faktoron keresztül vezették be a PD és az RR közti korrelációt (J. Frye és mtsai., 2000; J. H. Frye, 2000; Pykhtin, 2003; Tasche, 2004). Az ilyen modellek legnagyobb problémája az volt, hogy a vigaszára, illetve a nemteljesítés miatti veszteség csak egy szisztematikus kockázati tényező (determinisztikus) függvényei voltak, ami valljuk be, egy nagyon irrealisztikus feltevés. Ezért a kutatók bővebb modelleket alkottak meg annak érdekében, hogy a két kockázati tényező közti kapcsolatot pontosabban tudják modellezni.

A teoretikus modellek között alapvető szemléletbeli különbség van. Ez általában abból adódik, hogy a kutatók más-más típusú adatsorral, adatbázissal rendelkeznek, így különböző lépésekkel juthatnak el a keresett kapcsolati mérőszámokig. Például Witzany egy bank szemszögéből indul ki, amely egy nagy méretű, vállalati hitelek (kötvényekből) álló portfólióval rendelkezik (Witzany, 2011). E portfólió elemeit nem kereskedik minden nap, ezért a bank nem rendelkezik napi árfolyamokkal és loghozamokkal.

Ehelyett Witzany (ahogy a legtöbb pénzügyi intézet és szolgáltató) aggregált adatsorokkal dolgozik. Az adatsorok segítségével kiszámítja az egyes hitelkockázati csoportok nemteljesítési eseményeihez kapcsolódó valószínűségeit (PD) és a nemteljesítéskor fellépő veszteségek (LGD) mértékét, majd a felépített pénzügy-matematikai modell segítségével kiszámítja a korrelációs mérőszámokat.

Az alfejezet elején említett probléma feloldására Witzany a következő bővített modellt határozta meg (Witzany, 2011):

$$(3) \quad Y_1 = \sqrt{\rho_1} X_1 + \sqrt{1 - \rho_1} \xi_1$$

$$(4) \quad Y_2 = \sqrt{\rho_2} (\omega X_1 + \sqrt{1 - \omega^2} X_2) + \sqrt{1 - \rho_2} \xi_2$$

ahol X_1, X_2, ξ_1 és ξ_2 egymástól független, standard normális eloszlású valószínűségi változók. A nemteljesítés valószínűségét (PD) meghatározó Y_1 (*driver*) folyamatban ρ_1 az X_1 szisztematikus faktor főkomponenssúlya (*loading*). A nemteljesítés miatti veszteséget (LGD) meghatározó Y_2 folyamatban ρ_2 az összetett faktor főkomponenssúlya.

Witzany egyedi ötletét a (4) képletben lehet látni, ahol az Y_2 folyamat főkomponensét, a következő két szisztematikus faktor lineáris kombinációjaként definiálta:

- X_1 : a nemteljesítés valószínűségéhez (PD) tartozó szisztematikus faktor.
- X_2 : a nemteljesítés miatti veszteséghez (LGD) tartozó faktor, amely a behajtási időszak (*recovery process*) alatti gazdasági hatásokat foglalja magában.

Pontosabban fogalmazva, a PD szisztematikus faktora (X_1) és az LGD szisztematikus faktora (X_2) közötti kapcsolatot Witzany egy ω paraméterű Gauss kopula segítségével írta le. Az ω korrelációs paraméter ($\omega \in [-1; 1]$) értéke általában negatív, innen ered a *wrong-way risk* elnevezés is.

Vegyük észre, hogy az Y_1 és Y_2 driverek standard normális eloszlású valószínűségi változók, mert független (standard) normális eloszlású valószínűségi változók lineáris kombinációjaként lettek definiálva és a „megfelelő” súlyozásnak köszönhetően e változók várható értéke 0, a szórása pedig 1.

A pénzügyi szolgáltatók időegységenként (például havonta, negyedévente, évente) mérik a portfóliójukat (például fogyasztói hitelek, jelzáloghitelek) jellemző kockázati összetevők értékeit. Így a nemteljesítési ráták (valószínűségek) és a nemteljesítéskori várható veszteségek idősora a következőképp írható föl:

- $pd(t)$, ahol $t = 1; 2; \dots; T$
- $lgd(t)$, ahol $t = 1; 2; \dots; T$

Witzany elmélete alapján a PD és az LGD értékeit bizonyos mögöttes, látens tényezők befolyásolják. A cseh közgazdász a g és h függvényeken keresztül határozta meg a nemteljesítés valószínűségét (PD) és a nemteljesítéskori veszteség értékét (LGD) a piacon megfigyelhető szisztematikus faktorok alapján (a független, egyedi – idioszinkratikus – faktorok (ξ_1 és ξ_2) diverzifikálódnak):

- $x_1(t)$, ahol $t = 1; 2; \dots; T$
- $x_2(t)$, ahol $t = 1; 2; \dots; T$

A Witzany által megfogalmazott összefüggést az alábbiak szerint lehet formalizáltan leírni (fontos: az x_i tényezők a korábban meghatározott Y_i driverekkel egyeznek meg):

$$(5) \quad pd(t) = g(x_1(t)), \text{ ahol } t = 1; 2; \dots; T$$

$$(6) \quad lgd(t) = h\left(\omega x_1(t) + \sqrt{1 - \omega^2} x_2(t)\right), \text{ ahol } t = 1; 2; \dots; T$$

A cseh közgazdász egy korábbi cikkében definiálta a $g(x)$ függvényt, amely a Bazel II. szabályozás során alkalmazott Vasicek formulán alapszik (Witzany, 2008):

$$(7) \quad g(x) = P\left[\sqrt{\rho_1} x + \sqrt{1 - \rho_1} \xi_1 > y_D \mid x\right] = \Phi\left(\frac{\sqrt{\rho_1} x - y_D}{\sqrt{1 - \rho_1}}\right)$$

ahol $y_D = -\Phi^{-1}(PD_0)$. PD_0 a nemteljesítés várható valószínűségét (teljes portfóliót tekintve), Φ pedig a standard normális eloszlás eloszlásfüggvényét jelöli. A $g(x)$ függvényhez a PD_0 paraméter értékét a nemteljesítési valószínűségek idősorának átlagaként kaphatjuk meg, míg a ρ_1 korrelációs paraméter *maximum likelihood (ML)* becsléssel határozható meg.

Az $lgd(t)$ kiszámításához szükségünk van a nemteljesítés miatti veszteség számlaszintű (szerződés szintű) Q eloszlásfüggvényére. (Ennek a meghatározásával foglalkozom a szakdolgozatom másik felében.) Witzany a $h(t)$ függvényt az alábbi módon határozta meg:

$$(8) \quad h(x) = E\left[G(\sqrt{\rho_2} x + \sqrt{1 - \rho_2} \xi_2) \mid x\right] = \int_{-\infty}^{\infty} G(\sqrt{\rho_2} x + \sqrt{1 - \rho_2} w) \phi(w) dw$$

ahol $G = Q^{-1} \circ \Phi$. Összegezve a g függvény függ egy ismeretlen ρ_1 korrelációs paramétertől és a PD_0 nemteljesítés várható valószínűségétől, míg a h függvény az ismeretlen ρ_2 korrelációs paramétertől és a számlaszintű LGD eloszlásától (Q függvény) (Witzany, 2009).

Witzany a mögöttes két idősróról (x_1 és x_2) feltételezte, hogy standard normális eloszlású elemekből áll. Azaz e két idősort piacon megfigyelhető adatok transzformáltjai (standardizáltjai) alkotják.

A következő kézenfekvő feltételezés a stacionaritás lenne, azonban számos gazdasági idősor esetén autokorreláció lép föl, mert a mögöttes makroökonomia folyamatok jellemzően autokorreláltak. Annak érdekében, hogy ezt ki lehessen küszöbölni $ARMA(p, q)$ specifikációt tételezünk fel mindkét idősorra:

$$(9) \quad x_k(t) = \sum_{i=1}^p \alpha_{k,i} x_k(t-i) + \sum_{j=0}^q \beta_{k,j} u_k(t-j), \quad t = 1; 2; \dots; T \text{ és } k = 1; 2$$

ahol $\alpha_{k,i}$ és $\beta_{k,j}$ az $ARMA$ modell együtthatói és $u_k(t)$ hibafolyamat (innováció), amely független, standard normális eloszlású tagokból áll.

Witzany modelljének az esszenciája abban rejlik, hogy a megfigyelt PD és LGD idősorok alapján, az általa feltételezett struktúra mellett becslést tud adni a kulcsfontosságú paraméterekre (ρ_1 , ρ_2 és ω). Ahhoz, hogy modellt tudjunk illeszteni a megfigyelt adatsorokra

elengedhetetlen a hosszútávú nemteljesítés szintjének (y_D) és az LGD számlaszintű eloszlásfüggvényének (Q függvény) a meghatározása.

A modell végső célja a Bázeli II. szabályozásban meghatározott tőkekövetelmények kiszámítása, pontosabban a nem várt veszteség (UL : *unexpected loss*) becslése volt. Ez azért érdekes számunkra, mert a szabályozás implicit módon feltételezi, hogy a nemteljesítési esemény valószínűsége és a nemteljesítés miatti veszteség értékei korrelálatlanok, a modell pedig, pont abból indul ki, hogy ezen értékek igenis korrelálnak. Így kapunk két darab 1 éves, 99.9%-os UL becslést (egy korrelálatlant és egy korreláltat), melyet össze lehet hasonlítani.

Azonban nem csak aggregált adatokból lehet modellt építeni – ahogy azt a cseh közgazdász tette (Witzany, 2011) – hanem például napi szinten kereskedett portfóliók segítségével is. Az utóbbi eset azért fontos, mert ekkor felírható a portfólió napi szintű értékváltozása, amely segítségével a különböző kockázati tényezők – mint például a nemteljesítési esemény valószínűsége (PD) és a nemteljesítéskori veszteség (LGD) – is felírhatóak (az előző módszerhez képest más típusú összefüggésekkel). Ebből a perspektívából indult ki Bade, Rösch és Scheule elmélete (Bade és mtsai., 2011).

A szerzőhármás modellje jelentős részben a Merton modell feltételezésein és gondolatiságán alapszik (Merton, 1974). Így az általuk továbbfejlesztett modell szerint egy hitelfeltevő (általában vállalatok) akkor válik nemteljesítővé (*default-ossá*), ha az eszközeinek összértéke az adósság lejáratkori értéke alá esik.

Legyen V_{it} az i -edik hitelfeltevő vagy kötvénykibocsátó eszközérték-változásának természetes alapú logaritmus, azaz matematikai értelemben az eszközértékeinek a loghozama a t -edik időszakban. Ezt a loghozamot a következő faktormodellel írták föl a szerzők:

$$(10) \quad V_{it} = -(\beta_0 + \beta' x_{it}^V + \sqrt{\rho^V} F_t + \sqrt{1 - \rho^V} Z_{it}^V)$$

ahol $x_{it}^V = (x_{it1}^V, \dots, x_{itK}^V)'$ azaz x_{it}^V -t K darab piacon megfigyelhető, determinisztikus, cégspecifikus, iparág-specifikus vagy makrogazdasági kockázati tényező alkotja, amelyek befolyásolják a loghozamot. Továbbá $\beta = (\beta_1, \dots, \beta_K)'$ az x_{it}^V tényezőihez tartozó K darab érzékenység vektora, β_0 pedig egy konstans. F_t egy szisztematikus (valószínűségi változó) tényező, mely az összes eszköz hozamára hat együttesen (globálisan) a t -edik időszakban. Z_{it}^V pedig egyedi (idioszinkratikus) valószínűségi változó, amely csak az i -edik eszköz hozamát befolyásolja a t -edik időszakban. Az utóbbi két kockázati tényező súlyai rendre: $\sqrt{\rho^V}$ és $\sqrt{1 - \rho^V}$. Mindemellett a szerzők azt feltételezték, hogy az összes valószínűségi változó független, minden megfigyelt időpillanatban (Bade és mtsai., 2011).

Fontos kiemelni az x_{it}^V a hitelfelvevőre (az esetek többségében a vállalatokra) ható piaci hatásokat tartalmazza, az F_t a mindenkire (globálisan) ható, a piacot általánosan befolyásoló hatásokat írja le (ide tartozhat például az infláció, ami minden fogyasztóra és vállalatra hatással van), míg a Z_{it}^V egyedi vállalat-specifikus (maradék)változó. Így e modell alapján kijelenthető, hogy az x_{it}^V változó az egyéni kockázatokat egybefoglalva írja le, az F_t változó pedig rendszerszintű, szisztematikus kockázatokat foglalja magában.

A nemteljesítési esemény akkor következik be, ha az eszközérték-változás (hozam) átlép egy küszöbértéket. Általánosságban elmondható, hogy a bankok hitelportfóliójának többségénél nem figyelhető meg a portfólióban szereplő eszközök piaci értéke minden időpillanatban. Így a (10) egyenlet egy nem megfigyelhető folytonos változót (az eszközérték-változást) kapcsol össze egy megfigyelhető nemteljesítési (*default*) eseménnyel és a piacon megfigyelhető kockázati tényezőkkel (x_{it}^V).

Bade, Rösch és Scheule két fontos – a pénzügyi matematika területén általánosan elfogadott – feltevést fogalmazott meg. Az első – összhangban Robert Merton 1974-es modelljével – szerint F_t és Z_{it}^V egymástól független (standard) normális eloszlású valószínűségi változók, továbbá impliciten feltételezték a nem megfigyelhető V_{it} -ről is, hogy normális eloszlású egységnyi szórással. A második feltevésük pedig az volt, hogy a fent említett küszöbérték legyen 0, például β_0 segítségével.

Az általuk kidolgozott elmélet szerint a nemteljesítési (*default*) eseményeket meg lehet ragadni Bernoulli kísérlet segítségével (ebben az esetben D_{it} valószínűségi változó):

$$(11) \quad D_{it} = \begin{cases} 1, & \text{ha } i \text{ hitelfelvevő nemteljesítővé válik a } t. \text{ időpontban} \\ 0, & \text{egyébként} \end{cases}$$

Másrészt a nemteljesítési események megközelíthetőek – ahogy a Merton modell esetén is láthattuk – a vállalat eszközérték-változásával. Formalizáltan, ha V_{it} értéke 0 alá esik, a megfigyelhető x_{it}^V változók ismeretében, akkor *default* esemény következik be.

Így a Bernoulli-eloszlású valószínűségi változót felhasználva, az eszközérték változás segítségével definiálható a PD:

$$(12) \quad PD_{it} = P(D_{it} = 1 \mid x_{it}^V) = P(V_{it} < 0 \mid x_{it}^V) = \Phi(\beta_0 + \beta' x_{it}^V)$$

ahol Φ a standard normális eloszlás eloszlásfüggvényét jelöli. Az összefüggés jobb oldalából kiolvasható, hogy a PD értéke akkor lesz magasabb, ha az x_{it}^V értékei nőnek, azaz olyan piaci változások következnek be, amelyek hatására x_{it}^V értékei növekednek. Továbbá kiolvasható a (12) képletből, hogy a modell szerint csak az egyéni kockázatok hatnak (x_{it}^V) a nemteljesítés valószínűségének (PD) alakulására.

Mielőtt tovább haladnánk elengedhetetlen az eszköz (hozam) korrelációnak a definiálása. Két hitelfeltevő (vállalat) eszközértékváltozása közötti korrelációt ρ^V segítségével adhatjuk meg (a t-edik időszakban).

A Bazel II. szabályozás esetén alkalmazott IRB módszerrel összhangban a nemteljesítés feltételes valószínűsége (CPD: *conditional probability of default*) az $f_t = F_t$ szisztematikus kockázati tényező mellett az alábbiak szerint határozható meg:

$$(13) \quad CPD_{it}(f_t) = P(D_{it} = 1 | x_{it}^V, f_t) = P(V_{it} < 0 | x_{it}^V, f_t) = \Phi\left(\frac{\beta_0 + \beta' x_{it}^V + \sqrt{\rho^V} f_t}{\sqrt{1 - \rho^V}}\right)$$

ahol Φ a standard normális eloszlás eloszlásfüggvényét jelöli. A képletből kiolvasható, hogy a nemteljesítés feltételes valószínűsége (CPD) egy sokkal összetettebb kifejezés, mint a nemteljesítés valószínűsége (PD).

A CPD esetén már megjelenik a feltételes várható értékben az f_t is, mely a piacon szisztematikus módon fellépő hatásokat írja le, ezáltal a CPD nemcsak az egyéni, hanem a piaci kockázatokat is magában foglalja. Így a nemteljesítés feltételes valószínűsége nem csak az x_{it}^V értékváltozásaira érzékeny, hanem az f_t értékváltozására is (mindkét esetben akkor nő a CPD értéke, ha az adott változó értéke nő).

Az eddigiekben bemutatott Bade, Rösch és Scheule modelljének az első felét, amely a nemteljesítés valószínűségével (a PD-vel) foglalkozott, most pedig áttérek a vigaszrátákkal kapcsolatos elméletük bemutatására.

A szerzőhármas a vigaszráta modellezésére Pykhtin lineáris faktormodelljét vette alapul (Pykhtin, 2003). Erre építve alkották meg a vigaszráta mögötti Y_{it} (*driver*) folyamatot:

$$(14) \quad Y_{it} = \gamma_0 + \gamma' x_{it}^Y - \sqrt{\rho^Y} F_t + \sigma U_{it}$$

ahol $x_{it}^Y = (x_{it1}^Y, \dots, x_{itL}^Y)'$ azaz x_{it}^Y -t L darab piacon megfigyelhető, determinisztikus kockázati tényező alkotja, amelyek befolyásolják az a behajtási folyamatot (*recovery process*), ezáltal hatással vannak a vigaszráta értékére. Továbbá $\gamma = (\gamma_1, \dots, \gamma_L)'$ az x_{it}^Y tényezőkhöz tartozó érzékenységi vektora, γ_0 pedig egy konstans. F_t egy szisztematikus (valószínűségi változó) tényező, mely az összes eszköz hozamára hat együttesen (globálisan) a t-edik időszakban. E kockázati tényező súlya a modellben $\sqrt{\rho^Y}$. Végezetül pedig U_{it} egy standard normális eloszlású hibatermék, konstans σ súllyal. Az orosz származású fizikus-közgazdász elméletét követve, az Y_{it} (*driver*) folyamat segítségével meghatározták a vigaszráta értékét:

$$(15) \quad RR_{it} = \exp(Y_{it})$$

Ezek alapján leolvasható, hogy a vigaszráta esetén lognormális eloszlást tételeztek föl, melynek segítségével elkerülték, hogy negatív értékeket vegyen föl az RR.

Azonban a lognormális eloszlással van egy nagy „probléma”. Ebben az esetben a vigaszráta 0 és végtelen közötti értékeket vehet föl ($RR \in (0; \infty)$). A valóságban előfordulhat, hogy a RR 1-nél nagyobb értékkel rendelkezzon, de ez nagyon ritka. Emiatt kialakult az a gyakorlat, hogy a vigaszráta értékét a $[0; 1]$ intervallumra korlátozzák és e szűkített, zárt intervallumon modellezik azt.

Bade, Rösch és Scheule a vigaszrátákat az empirikus kutatásuk során kötvényárakból számolták ki, így ebben az esetben az RR (valós) értéke 0 és végtelen közé eshet. A vigaszráta értéke akkor és csakis akkor vehet föl 1-nél nagyobb értéket, ha a kötvény nemteljesítés utáni ára meghaladja a kötvény névértékét.

Közgazdaságtanilag Y_{it} úgy értelmezhető, mint a hitelező *default* esemény idején fennálló kitettségek (EAD) a potenciális hozama:

$$(16) \quad Y_{it} = \ln(RR_{it}) = \ln\left(\frac{LV_{it}}{EAD_{it}}\right) = \ln(LV_{it}) - \ln(EAD_{it})$$

ahol LV_{it} a vállalat (adós) eszközeinek likvidációs értéke.

A vállalat eszközértékváltozás folyamatát és a behajtási folyamatot (*recovery process*) – ezáltal a PD-t és az RR-t – a közös szisztematikus kockázatnak való kitettség kapcsolja össze. Ezt a (10) és a (14) egyenletben található F_t – standard normális eloszlású – valószínűségi változón keresztül figyelhetjük meg.

A szerzőhármas az U_{it} hibatagon keresztül kapcsolta össze a PD és az RR mögött meghúzódó egyéni faktorokat (Z_{it}^V és Z_{it}^Y). Úgy, hogy a U_{it} és Z_{it}^V valószínűségi változók korrelációja pont ρ^U legyen. Ezt a Cholesky-faktorizáció segítségével, az alábbiak szerint tudták megvalósítani:

$$(17) \quad U_{it} = \rho^U Z_{it}^V + \sqrt{1 - (\rho^U)^2} Z_{it}^Y$$

ahol Z_{it}^Y egy standard normális eloszlású valószínűségi változó, amely független Z_{it}^V -től. Vegyük észre, hogy ρ^U nagyon hasonló korrelációs paraméter, mint Witzany modellje esetén meghatározott ω paraméter, ahol a PD-t és az LGD-t meghatározó egyedi faktorokat kapcsolta össze általa.

A képletből látszik, hogy U_{it} (standard normális eloszlású valószínűségi változó.) szétválasztható két független, standard normális eloszlású valószínűségi változó (Z_{it}^V és Z_{it}^Y) összegére, amely számunkra megfelelő korrelációs struktúrában áll. E felbontás segítségével az vigaszráta mögötti folyamat felírható:

$$(18) \quad Y_{it} = \gamma_0 + \gamma' x_{it}^Y - \sqrt{\rho^Y} F_t + \sigma \rho^U Z_{it}^V + \sigma \sqrt{1 - (\rho^U)^2} Z_{it}^Y$$

Ezzel a lépéssel a két folyamatot külön-külön leíró egy-faktor modelleket kapcsolták össze, így összeségében a Witzany (2011) modelljéhez hasonló kétfaktoros korrelációs modellt alkottak meg. A szerzők a (18) egyenletből levezették a különböző változók közti korrelációs összefüggéseket. Ezt érdemes áttekinteni, mert ennek segítségével tisztább képet kaphatunk a változók között fennálló lineáris kapcsolatáról.

$$(19) \quad \rho^{\ln(RR)} = \frac{\rho^Y}{\rho^Y + \sigma^2} \quad \text{2 db log-recovery közötti korreláció}$$

$$(20) \quad \rho^{RR} = \frac{\exp(\rho^Y) - 1}{\exp(\rho^Y + \sigma^2) - 1} \quad \text{2 db vigaszráta közötti korreláció}$$

$$(21) \quad \rho^{VY} = \frac{\sqrt{\rho^V \rho^Y} - \sigma \rho^U \sqrt{1 - \rho^V}}{\sqrt{\rho^Y + \sigma^2}} \quad \text{eszköz értékváltozás és log-recovery között}$$

$$(22) \quad \rho^{V,RR} = \frac{\sqrt{\rho^V \rho^Y} - \sigma \rho^U \sqrt{1 - \rho^V}}{\sqrt{\exp(\rho^Y + \sigma^2) - 1}} \quad \text{eszköz értékváltozás és vigaszráta között}$$

A modell végső célja különböző veszteségek kiszámítása, például a várható veszteség (*EL: expected loss*) és a szisztematikus kockázati tényezőtől függő várható veszteség (*CEL*) kiszámítása. Ennek segítségével ki lehet kalkulálni a portfólió veszteségeloszlását, továbbá a portfólióhoz tartozó kockázatosított értéket (*VaR: Value at Risk*).

Mivel a várható veszteségeket a kitettséghoz arányosan mérik, ezért kiszámítható a vigaszráta várható értéke, illetve a vigaszráta feltételes várható értéke nemteljesítés (*default*) esetén. Itt nagyon fontos különbséget tenni a két mérőszám között, mert csak az utóbbit figyelhetjük meg a piacon. Ezért érdemes összehasonlítani az egyedi kötvények (értékpapírosított hitelek) vigaszrátáit, a nemteljesítés esetén tapasztalt RR értékekkel.

Vegyük észre, hogy a két fent bemutatott *wrong-way risk* modell, mely nagyjából ugyanazt az alapmodellt definiálja (változók közti kapcsolatok leírására), leginkább abban különböznek, hogy milyen típusú kiindulási adattal dolgoznak. Attól függően, hogy milyen perspektívából tekintettek a problémára, ahhoz igazították a folyamatokat leíró egyenleteket és összefüggéseket.

A két modell közötti legnagyobb (matematikai) hasonlóság a korrelációs struktúráknál található, mert az egyes korrelációs paraméterek nagyon hasonló változókat kapcsolnak össze. Az előző oldalon részletesen kifejtettem az ω és a ρ^U közötti párhuzamot. A legnagyobb (matematikai) különbség pedig az LGD, illetve RR mögött meghúzódó mögöttes eloszlásoknál található. Witzany esetében bármilyen eloszlást be lehet írni a G függvénybe (LGD eloszlásfüggvénye), melyet megfelelőnek gondolunk. Ezzel szemben Bade, Rösch és Scheule elmélete lognormális eloszlást tételezett föl a vigaszrátára (RR).

A következő táblázatban összegyűjtöttem a két modell gyakorlati alkalmazásának főbb hasonlóságait és különbségeit:

1. táblázat: *Wrong-way risk modellek összehasonlítása*

	<i>Witzany</i>	<i>Bade, Rösch, Scheule</i>
Milyen hiteltermékre vonatkozik a modell?	Bankhitel portfólió	Kockázatos, kereskedett kötvényportfólió
Milyen adatbázist használ fel?	Fedezetlen cseh lakossági hitelek portfóliója	Amerikai vállalati kötvények (Moody's adatbázisa)
Milyen típusú adatokkal dolgozik a modellben?	Aggregált adatok	Egyedi kötvények (napi szintű adatokkal)
Mire használ a modell mögöttes makroökonómiai faktorokat?	PD és LGD mögöttes faktorainak megalkotására	A mögöttes matematikai modell finomítására (modell bővítés)
Használ-e mögöttes eloszlást (RR-re vagy LGD-re)?	Igen (Q függvénnyel meghatározható)	Igen (lognormális eloszlást)
Mi a modell matematikai outputja?	ρ_1, ρ_2 és ω + illeszkedést segítő paraméterek	ρ^V, ρ^Y és ρ^U + érzékenységi paraméterek
Milyen módszerrel történik a becslés?	Maximum likelihood és bayesi MCMC	Maximum likelihood

2.3. A vigaszráta eloszlásának vizsgálata regressziós módszerekkel

Az előző alfejezetben már bemutattam a PD és az RR (illetve a PD és az LGD) közötti korrelációs kapcsolatot leíró pénzügymatematikai modelleket, ezért szakdolgozatom módszertani fejezetének második felét a vigaszráta és a nemteljesítés miatti veszteség bimodalitásának (multimodalitásának) vizsgálatára szenteltem.

Számos hitelkockázattal foglalkozó tanulmány fókuszál a vigaszráták bimodális alakjának vizsgálatára, azonban mai napig sincs általánosan elfogadott módszer, mellyel igazán hatékonyan lehet becsülni a vigaszráta eloszlását. Mivel az RR és az LGD eloszlását (illetve az eloszlást egyértelműen meghatározó sűrűségfüggvényt) nem lehet közvetlenül becsülni, ezért a kérdéssel foglalkozó kutatók ezt két lépésben, regressziós eljárások segítségével teszik meg.

Első lépésként a saját adatbázisuk alapján regressziós becslést adnak az RR vagy LGD értékekre, majd második lépésként, a becsült adatokra sűrűségfüggvényt illesztnek és ezt összehasonlítják az eredeti adatsorra illesztett sűrűségfüggvénnyel. Az elemzésem során én is ezt az eljárást fogom alkalmazni.

A különböző regressziós módszerek tulajdonságai és mögöttes matematikája számos információval látnak el bennünket a becsült adatokkal kapcsolatban – így például a magyarázó változók és a célváltozó közötti kapcsolatról, a becslés értékészletéről vagy épp a becslőfüggvény alakjáról –, melyek alapján releváns következtéseket vonhatunk le a valós adatokra. Természetesen a kutatók célja olyan regressziós eljárás keresése (a rendelkezésre álló módszerek közül a legjobb kiválasztása), mellyel a legpontosabb becslést tudják adni, hogy az arra illesztett sűrűségfüggvény (alakja és tulajdonságai) a lehető legjobban közelítse a valós sűrűségfüggvényt.

Ahogy azt a bevezetésben írtam, az egyik legnagyobb nehézség a vigaszráta, illetve a nemteljesítés miatti veszteség eloszlásának a becslésében, hogy az empirikus sűrűségfüggvényük általában 0 és 1 körül koncentrálódik, azaz bimodális tulajdonságú. Azonban néhány esetben még ennél is bonyolultabb alakú a sűrűségfüggvény. Így például Ye és Bellotti tanulmányában bemutatott adatsor esetén az LGD trimodális tulajdonságú (Ye & Bellotti, 2019).

A vigaszrátával foglalkozó regressziós módszereket két nagy csoportra lehet bontani: lineáris, illetve nemlineáris megközelítést alkalmazó modellek csoportjára. Az előbbieket viszonylag egyszerűen interpretálhatóak, a paramétereik (érzékenységek) könnyen értelmezhetőek, azonban számos hátrányuk van.

Példának vegyük a legegyszerűbb regressziós modellt, a legkisebb négyzetek módszerén alapuló *OLS* (*Ordinary Least Squares*) regressziót. Ekkor a vigaszrátára (vagy a nemteljesítés miatti veszteségre) felírható az alábbi modell (k darab magyarázó változó segítségével):

$$(23) \quad Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon = \mathbf{X}^T \boldsymbol{\beta} + \varepsilon$$

ahol Y a vigaszráta értéke az X_i -k a függő (magyarázó) változók, míg a β_i -k az illeszkedést segítő (k darab) paraméterek, továbbá ε az illeszkedést segítő hibatermék. Ezek alapján a becült vigaszráta értékeket az alábbiak szerint lehet kiszámolni:

$$(24) \quad \hat{Y} = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k = \mathbf{X}^T \boldsymbol{\beta} = Y - \varepsilon$$

Ezzel a megközelítéssel az gond, hogy a modell (lineáris) szerkezetéből adódóan \hat{Y} bármilyen értéket fölvehet a valós számok halmazán ($\hat{Y} \in (-\infty; \infty)$), ezzel szemben – az elméleti megközelítés szerint – a vigaszráta csak a 0 és 1 közötti intervallumon vehet föl értékeket ($RR \in (0; 1)$). Emiatt a becült értékekre illesztett sűrűségfüggvény értelmezési tartománya várhatóan nem korlátozódik kizárólag a $[0; 1]$ intervallumra.

A lineáris és a nemlineáris módszerek közti fő különbséget úgy lehet igazán jól megfogni, ha a nemlineáris kapcsolatot leíró egyenletet átranzformáljuk a g függvény segítségével úgy, hogy a függvényen belül megjelenjen a lineáris kifejezés:

$$(25) \quad Y = g(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$$

ahol g tetszőleges olyan invertálható függvény, amely teljesíti az alábbi feltételt:

$$(26) \quad g: \mathbb{R} \rightarrow [0; 1]$$

Ekkor garantáltan minden becült RR , illetve LGD érték a $[0; 1]$ intervallumon belül lesz. Így a nemlineáris módszerekből származó becslésekre illesztett sűrűségfüggvény értelmezési tartománya meg fog egyezni a valós adatokra illesztett sűrűségfüggvény értelmezési tartományával. Minden fenti feltételnek megfelelő g függvényt link függvénynek (*link function*) nevez a szakirodalom.

Ha vesszük a (25) egyenlet inverzét, akkor a jobb oldalon megjelenik a lineáris kifejezés, a bal oldalon, pedig valós vigaszráta vagy nemteljesítés miatti veszteség $(-\infty; \infty)$ értelmezett transzformáltja:

$$(27) \quad g^{-1}(Y) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

A fenti egyenlet esetén látszik igazán jól az általánosított lineáris modell (*GLM: generalised linear model*) megközelítés, mellyel a nemlineáris regressziós modelleket illesztik a gyakorlatban. Pontosabban az Y értékek transzformációja (g^{-1} függvény) segítségével a nemlineáris kapcsolat linearizálható. Az eljárás után kapott lineáris regresszió együttthatói hatékonyan és gyorsan megbecsülhetőek (az *OLS* módszer segítségével).

Mielőtt továbblépnék a különböző regressziós módszerek bemutatására, elengedhetetlen szót ejteni a regressziós módszerek teljesítményéről, előrejelző képességéről és összehasonlíthatóságáról. A modellek jóságát (teljesítmény-mérőszámok) nem csak mintán (adatbázison) belül érdemes tesztelni, hanem tanácsos valamilyen előrejelzést (*forecast*) is vizsgálni.

Ennek legegyszerűbb – és egyben évtizedek óta bevált – módja, ha az adatkészletet kettéosztjuk. Az első (nagyobb) rész a tanuló minta, amelyre a modellt illesztjük, ezáltal meg tudjuk határozni a modell paramétereit. A második (kisebb) rész a teszt minta, ahol a tanuló mintán kalibrált modellt teszteljük. A tanuló mintán összehasonlítjuk a becsült és az eredeti értékeket, ezáltal meg tudjuk határozni a modell mintán kívüli teljesítőképességét. Ezt az eljárást hívják egyszerű validációnak.

A kutatók a mintán kívüli előrejelzési teljesítményen (*forecast performance*) alapuló következtetéseket megbízhatóbbnak tartják, mint a mintán belüli teljesítményen alapuló következtetéseket, amelyek érzékenyebbek lehetnek a kiugró értékekre. Továbbá a mintán kívüli előrejelzések jobban tükrözik a kutatók számára a „valós időben” érkező információk kiszámíthatatlanságát (mivel nem látunk a jövőbe).

Az egyszerű validációt többször elvégezve – az adatsokaságot véletlenszerűen, de ugyanolyan arányban felosztva – nagyon különböző teljesítmény-mérőszámokat kaphatunk. Az esetleges nagy szórás elkerülésére érdekében keresztvalidációt (*cross-validation*) kezdtek alkalmazni a kutatók.

A leggyakrabban használt módszer a k -szoros keresztvalidáció (*k-fold cross-validation*). Az eljárás lényege, hogy az adatbázist felosztjuk k egyenlő méretű részre (szeletre) és k modellt tanítunk, mindegyiket $k-1$ részen. Minden modell esetén becslést (előrejelzést) végzünk a maradék egy szeleten és vizsgáljuk e modellek jóságát. Az így kapott k darab teljesítmény-mérőszám átlagát véve egy pontosabb értéket kapunk, ami alapján meg tudjuk határozni, hogy mennyire jó az adott modell. Tehát a keresztvalidáció az egyszerű validációnál precízebb, pontosabb módszer.

Többször említettem már, hogy a regressziós módszerek jóságát különböző teljesítménymérőszámokkal mérhetjük, melyek közül be fogok néhányat mutatni (Draper & Smith, 1998):

- Többszörös determinációs együttható (R^2)
- RMSE (*Root Mean Squared Error*)
- MAE (*Mean Absolute Error*)
- MAPE (*Mean Absolute Percentage Error*)

A regressziós modellek jóságának legszélesebb körben használt mutatója az R^2 . Megmutatja, hogy a függő változók (varianciája) együtt hány százalékban magyarázzák az eredményváltozó varianciáját. Az R^2 0 és 1 között vehet föl értéket, legrosszabb esetben nullát, míg ideális esetben egyet.

Fontos kiemelni, hogy e mutatószám érzékeny a magyarázó változók számára. Adott modell esetén a teljesítménymérőszámok nagyságából jelentős következtetéseket nem érdemes levonni (alacsony vagy közepes R^2 -tel rendelkező modellek gyakran használnak a való életben), elkerülve például az egymást magyarázó változók (*confounding variable*) modellbe építését.

Az *RMSE*: a becült (előrejelzett) és megfigyelt értékek közötti átlagos négyzetes különbség négyzetgyöke. 0 és végtelen közt vehet föl értékeket, ideális esetben nullát, míg legrosszabb esetben bármilyen nagy szám is lehet. Az *RMSE* egy igen elterjedt mérőszám, azonban sokszor megfogalmazzák vele szemben kritikaként, hogy négyzetes eltérést tartalmaz, amely felnagyítja a vizsgált eltéréseket, ami pénzügyi adatok becslésének vizsgálatokor problémákat okozhat.

A *MAE*: a becült (előrejelzett) és megfigyelt értékek közötti átlagos abszolút különbség. 0 és végtelen közt vehet föl értékeket, ideális esetben nullát, míg legrosszabb esetben bármilyen nagy szám is lehet. A *MAE* szintén nagyon népszerű mérőszám, melyet legtöbbször az *RMSE*-vel hasonlítanak össze. Ez a két legpontosabb és leggyakrabban használt teljesítménymérőszám.

A *MAPE* a becült (előrejelzett) és megfigyelt értékek közötti átlagos abszolút százalékos különbség. 0 és egy között vehet föl értékeket, ideális esetben nullát (0%), míg legrosszabb esetben egyet (100%). A *MAPE* nagy ismertségnek örvend, mivel az egyik leggyakoribb mérőszám az előrejelzések hatékonyságának vizsgálatára, elsősorban skálafüggetlensége és értelmezhetősége miatt. A *MAPE* azonban 0-hoz közeli értékek becslése esetén sokszor értelmezhetetlen (vagy végtelen) értékeket vesz föl, emiatt használata nem mindig ajánlott (S.

Kim & Kim, 2016). Így például RR, illetve LGD értékek becslésénél előnytelen ezt a mutatószámot használni.

A szakemberek körében a mai napig vita tárgyát képezi, hogy melyik a legjobb, legpontosabb mutatószám a modellteljesítmény és az előrejelzések jóságának a mérésére (Chai & Draxler, 2014). A szakdolgozatom szintjén úgy érzem az R^2 , az *RMSE* és a *MAE* egyaránt releváns (megfelelő) teljesítmény-mérőszámok. A segítségükkel összehasonlíthatóak – az ugyanolyan magyarázó változókat tartalmazó – regressziós módszerek.

A kutatók az elmúlt évtizedben olyan regressziós modelleket (módszereket) tanulmányoztak, amelyek kifejezetten a bimodalitás (illetve multimodalitás) problémájával foglalkoztak. Bellotti és Crook például *OLS* regressziót, Tobit, illetve döntési fa modellt (*DTM: Decision Tree Model*), valamint béta és frakcionált logit transzformációt használtak a nemteljesítés miatti veszteség modellezésére. Adatkészletük 55 000 darab nemteljesítő hitelkártya adatain alapult, amelyek 1999 és 2005 között váltak *default*-ossá az Egyesült Királyságban. Meglepő módon a felsorolt módszerek közül az *OLS* regresszió (legkisebb négyzetek módszere) teljesített legjobban és adta a legpontosabb előrejelzést makroökonómiai változók esetén (Bellotti & Crook, 2012).

Calabrese egy kevert, diszkrét-folytonos megközelítést vezetett be a vigaszráta becslésére, ahol a 0 és 1 határpontokat Bernoulli eloszlású valószínűségi változókkal, a 0 és 1 közötti nyílt intervallumon található értékeket, pedig folytonos, béta eloszlású valószínűségi változókkal modellezte. A kevert eloszlás segítségével a modell kezelni tudta az RR értékeinek a két határpont környékén tapasztalható sűrűsödését. Calabrese az Olasz Központi Bank (Banca d'Italia) 1985 és 1999 közötti hitelállományán tesztelte a modell pontosságát és előrejelző képességét (Calabrese, 2014).

A modelljét összehasonlította Papke és Wooldridge frakcionális vagy tört-válaszmodelljével (*fractional response model*), amelyet széles körben használnak vigaszráták modellezésére (Papke & Wooldridge, 1996). Calabrese az alábbi három gyakran használt link függvényt vizsgálta behatóbban:

- logisztikus (*logit*)
- log–log (*loglog*)
- kiegészítő log–log (*cloglog: complementary log–log*)

A fő különbség e három linkfüggvény között az, hogy a logit link függvény szimmetrikus, a log–log jobbra ferde, a kiegészítő log–log pedig balra ferde.

Az olasz kutató elemzéséből kiderült, hogy az általa javasolt modell valamennyivel jobb (előrejelzési) teljesítményt nyújt, mint a frakcionális válaszmodellek. Az általa kidolgozott modell révén becsült értékek rendelkeznek a vigaszráták egyedi (elméleti) tulajdonságaival, hasonlóan a frakcionális válaszmodellekhez.

Qi és Zhao hat modellezési módszert hasonlított össze, melyekkel a nemteljesítéskori veszteséget (LGD) modellezték a *Moody's Ultimate Recovery Database* nevű adatbázisa alapján, ami az Egyesült Államokban 1985 és 2008 között nemteljesítő, 3 751 hitel és kötvény adatait tartalmazta (Qi & Zhao, 2011).

Az elemzésük alapján arra a következtetésre jutottak, hogy a nemparametrikus módszerek – regressziós fa (*NRT: Nonparametric Regression Tree*) és neurális hálózat – jobban teljesítenek, mint a parametrikus módszerek – *OLS* regresszió, törtválaszregresszió (*FRM: Fractional Response Model*) és inverz Gauss regresszió (*IGR Inverse Gaussian Regression*) – mind a mintán belül, mind azon kívül, ha a túlillesztést megfelelően szabályozzuk.

A parametrikus módszerek közül a frakcionált válaszregresszió (*FRR*) rendelkezett a legjobb előrejelző képességgel a teljesítmény-mérőszámok alapján, nem sokkal megelőzve az *OLS* regressziót. Fontos kiemelni, hogy a parametrikus módszerek közé tartozó transzformációs (általánosított lineáris regresszió alapuló) módszerek – inverz Gauss regresszió (*IGR*) és béta-transzformált inverz Gauss regresszió (*IGR-BT: Inverse Gaussian Regression with Beta Transformation*) – teljesítménye nagyon érzékeny a transzformáció során használt ε paraméter nagyságára. Azonban optimális ε (nehéz meghatározni) megválasztása mellett nagyjából megegyező teljesítményt nyújtanak az *OLS* regresszióval.

Túl kicsi vagy túl nagy ε paraméter választása esetén a modell előrejelző képessége elég gyenge, ezzel szemben kis ε mellett a becsült LGD-k sűrűségfüggvénye igen erőteljes bimodalitást mutatott, összhangban a bevezetésben bemutatott elmélettel. Ezek alapján a kutatópáros azt a tanulságot vonta le, hogy az erős bimodális-mintázatot produkáló modellek sem feltétlenül rendelkeznek jó modellilleszkedéssel és pontos LGD előrejelzésekkel. Vagyis olyan regressziós eljárásokat kell keresni, melyek viszonylag jó előrejelző képességgel rendelkeznek, mindamelllett, hogy a becsült sűrűségfüggvény (aránylag) jól illeszkedjen az eredeti adatokra.

Loterman és munkatársai egy nagyszabású, módszertani kutatás (*benchmark study*) keretein belül összehasonlították a nemteljesítés miatti veszteség (LGD) eloszlásával foglalkozó regressziós modelleket. Összesen huszonnégy különböző modellt vizsgáltak hat nemzetközi banktól kapott adatkészlet alapján.

Ezek közé tartoztak az egyszerű (*one-stage*) modellek, például az egyszerű legkisebb négyzetek regressziós módszere (*OLS regression*), béta regresszió (*beta regression*), robusztus regresszió (*robust regression*), „ridge” regresszió (*Tikhonov regularization* vagy *ridge regression*), regressziós spline-ok (pl *MARS modellek: Multivariate Adaptive Regression Splines*), neurális hálózatok (*ANN: Artificial Neural Networks*), támaszvektor-gépek (*SVM: Support Vector Machine*) és regressziós fákkal (*regression tree*) felépített modellek, valamint több technikát kombináló összetett vagy kevert (*two-stage*) modellek.

Az elemzésükből kiderült, hogy egyik modell sem magyarázta igazán jól az LGD varianciáját (az R^2 értéke 4% és 43% között mozgott), ezáltal egyik modell előrejelző képessége sem volt igazán erős – erre a megfigyelésre jutott a legtöbb a témában írt publikáció – bár ez várható volt annak fényében, hogy mennyire összetett a nemteljesítés miatti veszteség (LGD) folyamata.

Továbbá arra a következtetésre jutottak, hogy a nemlineáris, illetve bizonyos gépi tanulási módszerek (*machine learning*), például a neurális hálózatok és támaszvektor-gépek jobban teljesítenek, mint a hagyományos lineáris módszerek. Ezenkívül a lineáris és nemlineáris technikák kombinációjával felépített összetett (kevert) modellek hasonlóan jó prediktív képességgel rendelkeznek (a rendelkezésre álló modellek között), azzal a további előnnyel, hogy van egy könnyedén értelmezhető lineáris modellkomponensük (Loterman és mtsai., 2012).

Krüger és Rösch alkalmazta először a kvantilis regressziót (*QR: Quantile Regression*) nemteljesítés miatti veszteség (LGD) modellezésére. A *QR* gyökeresen különbözik az eddig bemutatott módszerektől, mert az eljárás során nem a feltételes átlagot becsüli a magyarázó változók alapján, hanem a feltételes kvantilisekre – általában a mediánra – ad becslést. Pontosabban a kvantilis regresszió egy célváltozó változó (esetünkben az LGD) feltételes kvantiliseit a magyarázó változók lineáris függvényeként fejezi ki. Tehát a *QR* a lineáris regresszió kiterjesztésének, bővebb változatának is tekinthető.

A kvantilis regresszió két nagy előnye van a hagyományos regressziós modellekkel szemben. Egyrészt az eljárás elvégzéséhez nem szükséges semmilyen mögöttes eloszlás feltételezése, másrészt a módszer ellenálló a kiugró értékek hatásaira (ezáltal ez a *QR* egy robusztus módszer).

Esetünkben a módszer segítségével a középső kvantilisek értékei jól magyarázhatók a megfigyelhető kovariánsok alapján. Ezzel szemben arra az érdekes következtetésre jutottak a kutatók, hogy a farokeseményeket – például a rendkívül magas (1-hez közeli) LGD értékeket –

leginkább a nem megfigyelhető, véletlenszerű események vezérlik, így ezeket nem lehet a konvencionális regressziós módszerekkel hatékonyan magyarázni (Krüger & Rösch, 2017).

Az általuk implementált módszernek köszönhetően behatóbban megismerhetjük a vigaszráta és a nemteljesítés miatti veszteség viselkedését. A tanulmány további érdekessége, hogy a teljes LGD eloszlásra (sűrűségfüggvényre) vonatkozó becslések validálásához (értékeléséhez) alternatív illeszkedési (*goodness of fit*) mérőszámokat javasoltak.

Őket pár évvel megelőzve Siao, Hwang és Chu alkalmazta a logisztikus kvantilis regressziót (*LQR: Logistic Quantile Regression*) a vigaszráták elemzésére. A *Moody's Ultimate Recovery Database* nevű adatbázisa alapján megmutatták, hogy a becsült minta különböző részeiben (partícióiban) a vigaszráta értékek eltérő eloszlásúak (nem illeszthető rá egyben egy nagy eloszlás), ezért minden ilyen részlethez egy hibaminimalizáló kvantilispont került meghatározásra. Az *LQR* segítségével kiszámították az egyes kvantilis pontokhoz tartozó kvantilis regressziós görbéket (*quantile regression curves*), ezáltal pontosabb képet kapva a vigaszráták eloszlásának viselkedéséről. Ezek alapján beazonosítható, hogyan kapcsolódnak a magyarázó változók a vigaszráták mögöttes feltételes eloszlásához, különösen akkor, ha a mögöttes eloszlások erőteljesen ferdek.

Empirikus eredményeik azt mutatják, hogy a kilenc vizsgált modell (ezek közé tartoztak a korábban bemutatott modellek közül az *IGR*, az *IGR-BT*, a *DTM*, az *NRT* és Calabrese kevert modellje) közül a hibaminimalizáló kvantilisponttal rendelkező *LQR* modell rendelkezik a legjobb (az RMSE teljesítmény-mérőszám alapján) és legrobusztusabb előrejelzési teljesítménnyel. Ezáltal a logisztikus kvantilis regresszió egy hatékony alternatíva a vigaszráták modellezésére és tanulmányozására (Siao és mtsai., 2015).

A számos pozitív tulajdonsága mellett a kvantilis regresszióval kapcsolatban ki kell emelnem két negatívumot. Egyrészt a módszer adatigényes, mivel kvantilisenként becsül értékeket, ebből következik a második probléma, hogy az eljárás igen számításigényes a többi módszerhez képest.

Keller, Nagl és Rösch kibővítette a lineáris kvantilis regressziót (*LQR+*) egy neurális hálózati struktúrával annak érdekében, hogy rugalmasabban tudják közelíteni a nemteljesítés miatti veszteség eloszlását az egyes kvantilisekben. E fejlesztés lehetővé tette a kutatók számára, hogy bármilyen interakciót és nemlineáris hatást modellezni tudjanak anélkül, hogy a pontos formát előzetesen definiálniuk kellene. Az egyes kvantilisekben az előrejelzések pontossága akár 30%-kal is nagyobb lehet, mint az összehasonlításhoz használt modellek esetén, különösen a magasabb kvantiliseknél, amelyek a legfontosabbak a hitelkockázatban az esetleges nagy veszteségek miatt (Kellner és mtsai., 2021).

A következő táblázatban összefoglalom a bemutatott modelleket és a kutatók által elért eredményeket:

2. táblázat: Az *RR*-t és a *LGD*-t becslő regressziós modellek összehasonlítása

<i>Szerző(k)</i>	<i>Modell(ek)</i>	<i>Eredmények</i>
Bellotti és Crook	<i>OLS</i> regresszió, Tobit és döntési fa modellek, béta és fracionált logit transzformáció	Ezek közül az <i>OLS</i> regresszió a legpontosabb az <i>LGD</i> értékek előrejelzésére
Calabrese	Kevert, diszkrét-folytonos regressziós modell (Bernoulli + béta eloszlás)	Jobb előrejelző képesség, mint <i>OLS</i> vagy <i>FRM</i> regressziók, a kevert modell sem tudja reprodukálni a multimodalitást
Qi és Zhao	<i>OLS</i> regresszió, <i>FRM</i> , <i>IGR</i> , <i>IGR-BT</i> , regressziós fa és neurális háló	Pontosabb előrejelzés nemparaméteres módszerekkel, erős bimodális jelleg \neq jó modellilleszkedés
Loterman	24 modell összehasonlítása (<i>benchmark study</i>)	Modellek pontatlansága (R^2 értéke 4% és 43% között) nemlineáris módszerek jobbak (SVM és neurális háló) kevert modellek extra előnyei
Krüger és Rösch	Kvantilis regresszió (<i>QR</i>)	A farokeseményeket véletlenszerű események vezérlik, hitelkockázati mérőszámok pontosabb becslése (<i>pl. UL és Downturn LGD</i>)
Siao, Hwang és Chu	Logisztikus kvantilis regresszió (<i>LQR</i>)	A minta különböző partícióban eltérő a vigaszráták eloszlása, jobb előrejelző képesség, mint az eddigi modelleknek
Keller, Nagl és Rösch	Lineáris kvantilis regresszió (<i>LQR</i>) egy neurális hálózati struktúrával kiegészítve	A modell flexibilisebb, mint az addigi módszerek \Rightarrow pontosabb előrejelzés

2.4. Két módszertan összehasonlítása

A hitelkockázaton belül a vigaszrátákkal és a nemteljesítés miatti veszteségekkel foglalkozó szakemberek figyelmét leginkább a következő két jelenség köti le:

- I. az RR (illetve LGD) és a PD közt fennálló (lineáris korrelációs) kapcsolata
- II. a vigaszráták bimodális (multimodalitás) tulajdonsága

Az előző két alfejezetben bemutattam a két jelenséghez tartozó modellezési módszereket, most pedig az alábbi táblázat segítségével hasonlítom össze a két kutatási dimenzióhoz tartozó módszertant:

3. táblázat: A modellezési módszertan összehasonlítása

	<i>Wrong-way risk modellek</i>	<i>Regressziós modellek (bimodalitás)</i>
Mit vizsgál a modell?	RR vagy LGD és PD közti korrelációt	RR vagy LGD értékek mögöttes eloszlását, (bi- vagy multimodalitás)
Milyen matematikai modell áll a módszer mögött?	Két faktoros modell (pénzügyi matematika)	Regressziós modell (matematikai statisztika)
Milyen típusú adatokra van szükség?	Aggregált adatok vagy napi árfolyamok + mögöttes faktorok	Az egyedi hitelek leíró szocioökonometriai adatok
Lehet-e csak aggregált adatokkal használni a modellt?	Igen	Nem, egyedi adatok szükségesek
Milyen konkrét adatokra van szükség a modellezéshez?	PD, RR vagy LGD és látens változók	RR vagy LGD és egyedi magyarázó változók
Módszertan nehézsége?	Bonyolult becslés: maximum likelihood fv. és MCMC szimuláció	Két lépéses becslés: regresszió és sűrűségfüggvény illesztés

Első ránézésre nagyon különbözőnek tűnhet a két jelenség modellezése, azonban mindkét vizsgált dimenzió nagyon is összefüggő. E kapcsolatra mutatott rá Daniel Rösch. A német közgazdász professzor társszerzője az általam másodikként ismertetett, *wrong-way risk* modellt definiáló cikknek (Bade és mtsai., 2011) és szintén társszerzője két kvantilis regressziót alkalmazó tanulmánynak, melyet összefoglaltam az előző alfejezetben (Krüger & Rösch, 2017; Kellner és mtsai., 2021). További kapcsolatot jelent, ha belegondolunk milyen módszerekkel találkoztunk a képzésünk során, akkor megállapíthatjuk, hogy mindkét módszertan beleillik az elsajátított kvantitatív pénzügyi eszköztárba.

Véleményem szerint a legnagyobb különbség a két terület (jelenség) modellezése között, hogy milyen típusú adatokra van szükség. Míg az első terület esetén a PD és RR vagy PD és LGD értékeken kívül csak néhány látens makroökonómiai változóra van szükség, amelyekkel a gazdasági környezetet lehet leírni, addig a második terület esetén – a regressziós eljárások elvégzéséhez – szükség van különböző piaci és társadalmi jelenségeket, illetve egyedi tulajdonságokat leíró (szocioökonometriai) változókra is.

Az általam bemutatott modellek jelentős limitációkkal rendelkeznek, így jelenleg még nem kapcsolható össze a két modellezési terület, azonban úgy vélem, hogy a vigaszrátákkal foglalkozó modellek fejlődése a közeljövőben eljut arra a szintre, amikor e két dimenziót össze lehet kötni a modellen belül, ezáltal pontosabb és objektívebb előrejelzéseket és becsléseket lehet tenni.

A következő két fejezetben a gyakorlati kérdések elemzésével kívánok foglalkozni. Először bemutatom a rendelkezésemre álló adatbázisokat és a hozzájuk kapcsolódó adatokat (3. fejezet), majd a gyakorlatban is bemutatom hogyan működnek a modellek (4. fejezet).

3. Adatbázisok bemutatása

Az empirikus kutatásomhoz szükséges adatbázist két nagy adatforrás összekapcsolásával hoztam létre. Ennek legjelentősebb eleme a nyilvánosan elérhető Bondora (*Public Reports*) adatbázis hitelekkel kapcsolatos adatkészlete (*Loan dataset*). Kiegészítő jelleggel pedig felhasználtam az Eurostat három – az egy főre jutó GDP-vel, a munkanélküliségi rátával és az inflációval (harmonizált fogyasztói árindex alapján) kapcsolatos – adatkészletét.

Kutatásom során nyilvánosan elérhető adatokkal akartam dolgozni, ezért választottam ezt a két adatbázist. Annak, hogy egy *peer-to-peer lending-re* építő hitelezési szolgáltatás adataival dolgozom két előnye is van. Egyrészt ezen a területen rengeteg, nyilvánosan elérhető, számlaszintű adat van – szemben a kereskedelmi bankok hitelportfólióival – amelyeket fel tudok használni a modellezésemhez. Másrészt egy nagyon érdekes és fejlődő piaci szegmensről van szó, amelyet érdemes vizsgálni.

3.1. A Bondora adatbázis bemutatása

A Bondora egy Észtországbán 2008-ban alapított, de mára Európa több országban tevékenykedő hitelintézet, amely jelentős hitelportfólióval rendelkezik Észtországbán, Finnországbán és Spanyolországban. Manapság a pénzintézet már több befektetési és adatelemzési (statisztikai) szolgáltatást nyújt.

Az adatbázisom megalkotásához a hatalmas *Loan data* adatkészletet vettem igénybe, amely 112 attribútum segítségével mutatja be a Bondora hitelállományát (letöltéskor 221 341 hitelről tartalmazott adatokat). Az adatkészletben szereplő változók közül az alábbiakat használtam föl:

- *Country*
- *LoanDate*
- *DefaultDate*
- *MaturityDate_Original*
- *ProbabilityOfDefault*
- *LossGivenDefault*
- *EADI*

A *Country* változó megadja a hitelfelvevő lakóhelyét (országát). A *LoanDate* változó tartalmazza a hitelszerződés életbelépésének dátumát. A *DefaultDate* mutatja meg az adott hitel esetén a nemteljesítés időpontját (és a behajtási folyamat kezdetét), amennyiben ilyen esemény nem történt, akkor a dátum helyett az „NA” jelölés szerepel. A *MaturityDate_Original* változó tartalmazza a hitel szerződés szerződéses lejáratának időpontját (dátumát). A *ProbabilityOfDefault (PD)* változó a nemteljesítés valószínűségét mutatja egy éves időhorizonton belül (*credit-ratings*ek alapján számolva). A *LossGivenDefault (LGD)* változó pedig megadja a nemteljesítés időpontjában fennálló kitétség százalékos arányát, amelyet a hitelintézet várhatóan elveszít. E változó esetében is „NA” jelölés szerepel, ha az adott hitel még nem vált *default-ossá*, azaz nem következett be nemteljesítési esemény. Végezetül pedig az *EADI* változó tartalmazza a nemteljesítéskori kitétség (nemteljesítéskor fennálló tőketartozás) mértékét. Az utóbbi változót elővigyázatosságból válogattam be az adatbázisba, mert a segítségével ki lehet szűrni az olyan hiteket, amelyek esetében hiányos az adatközlés vagy azon eseteket, ahol még nem következett be nemteljesítési esemény.

A kezdeti adatbázisomhoz még négy változót határoztam meg a már meglévők alapján. Elsőként kiszámoltam a vigaszráták értékét a (2) egyenlet alapján, az LGD változó segítségével. Ezt követően negyedévekbe soroltam a *DefaultDate*, *MaturityDate_Original* és *LossGivenDefault* változókat, annak érdekében, hogy jobban követhető legyen az aktív hitelállomány változása. Továbbá követhetővé válik, hogy az egyes negyedévekben hány hitel válik nemteljesítővé. Ezáltal kiszámolható egy valós PD, vagyis az adott negyedévben a hitelek mekkora része válik nemteljesítővé.

A Bondora adatbázis 2009-től (a legkorábbi nemteljesítés 2009. április 16-ai) kezdődően napjainkig (az adatbázis letöltéséig) tartalmaz nemteljesítéssel kapcsolatos adatokat. A kezdeti mérési és becslési (modell kalibrálási) nehézségek miatt azonban sokáig pontatlan értékek kerültek az adatbázis oszlopaiba (például sok esetben 0 került a nemteljesítés valószínűsége [PD] oszlopába, ami teljesen irreális). Így az adatsor beható vizsgálata után úgy döntöttem, hogy kiszűrök minden olyan hitelt, amely esetében a PD változó nullát vesz föl, illetve azon hiteleket, melyek 2016 előtt visszafizetésre kerültek vagy nemteljesítővé váltak.

Továbbá kiszűrtem 2022-ben kötött új hiteleket, illetve az ebben az évben nemteljesítővé váló hiteleket, mert ezek nagyon friss adatok, javarészt még a behajtási folyamat sem játszódhatott le, ezért például az LGD értékek sem lehetnek pontosak. E két lépés hatására az adatbázis 201 300 eleműre szűkült, azonban ezek az adatok már hiánytalanok és ezáltal föl lehet használni az elemzéshez.

Az adatbázis értékei a következő négy országból származnak: Észtország, Finnország, Spanyolország és Szlovákia. Ezek közgazdaságtani értelemben igen heterogén országok, ezért a későbbi modellezés és elemzés során érdekes lesz, hogy mennyire különböznek például a PD és LGD közötti korrelációs (becsült)mérőszámok az egyes esetekben.

Az adatbázishoz tartozó dokumentációból kiderül, hogy az új piacokon, ahol korlátozott tapasztalatokkal rendelkezik a Bondora szolgáltató, nem tud pontos becsléseket adni és jól működő modelleket felállítani. Szlovákia egy új piac a Bondora számára és az adatelemzés során kiderült, hogy itt még egyértelműen a modell kalibrációs időszaka zajlik (kis számú adat van és azoknál is az alapértelmezett értékek szerepelnek), illetve kevés nemteljesítés történt, ezért a szlovák adatok nem hasznosíthatók érdemi elemzésre. Így ezen adatokat is kiszűrtem az adatbázisból.

3.2. Az Eurostat adatbázis bemutatása

Modellezéshez a Witzany által meghatározott korrelációs struktúrát használom föl (Witzany, 2011). E modellhez látens faktorokra is szükség van, amelyek az adott ország vagy régió gazdasági folyamatait írják le. Az ehhez szükséges – egy főre jutó bruttó hazai termékkel, munkanélküliségi rátával és inflációval kapcsolatos – adatokat az Eurostat megfelelő adatbázisaiból töltöttem le.

A bruttó hazai termék (*GDP: gross domestic product*) annak a hozzáadott értéknek a standard mérőszáma, amely egy országban adott időszak alatt az áruk és szolgáltatások előállítására keletkezik. A GDP a legelterjedtebb mutatószám a gazdasági aktivitás (adott ország gazdasági erejének) mérésére, azonban nem megfelelő mérőszám az emberek általános anyagi jólétének leírására. Ennek ellenére – talán az egyszerűségének és közérthetőségének köszönhetően – napjainkban is ez a leggyakrabban használt mutatószám a gazdaságok általános teljesítményének leírására. Ennél egy fokkal pontosabb képet mutat összehasonlítás esetén az *egy főre jutó bruttó hazai termék (GDP/fő)*, ezért én is ezt a mérőszámot használtam első látens változóként a három ország általános gazdasági helyzetének leírására.

Az ehhez szükséges adatokat az Eurostat: *Main GDP aggregates per capita* nevű adatbázisából (melynek az egyedi online adatkódja: *NAMQ_10_PC*) töltöttem le, ahol 2000-től negyedéves bontásban, kiigazítatlanul (sem szezonálisan, sem naptárhatással nincs kiigazítva), euróban, piaci értéken számítva szerepelnek az egy főre jutó bruttó hazai össztermék (GDP/fő) értékei.

A munkanélküliség a gazdaság egészségének kulcsfontosságú szocioökonómiai mérőszáma. A munkanélküliség olyan helyzetre utal, amikor egy aktívan állást kereső személy nem tud munkát találni (aki már nem keres aktívan munkát, az a gazdaságilag inaktív csoportjába tartozik, mint például a diákok és a nyugdíjasok). A munkanélküliség leggyakoribb mérőszáma a *munkanélküliségi ráta (index)*, amely a munkanélküliek száma osztva a teljes munkaerő-állomány (gazdaságilag aktívak = foglalkoztatottak + munkanélküliek) tagjainak számával.

Munkanélküliségről akkor beszélhetünk, ha a dolgozni akaró munkavállalók nem tudnak munkát találni, ami csökkenti a gazdasági teljesítményt. A magas munkanélküliségi ráta a gazdasági nehézségeket, a rendkívül alacsony munkanélküliségi ráta a gazdaság túlfűtöttségét jelezheti, amely szintén problémákhoz vezethet. Ezért a munkanélküliségi ráta egy igazán fontos és gyakorta használt mérőszáma az adott gazdaság teljesítőképességének. Emiatt választottam második látens faktoromnak a munkanélküliségi rátát.

Az ehhez szükséges adatokat az Eurostat *Unemployment rate - quarterly data, seasonally adjusted* nevű adatbázisából (melynek az egyedi online adatkódja: *TIPSUN30*) töltöttem le, ahol 2009-től negyedéves bontásban, szezonálisan kiigazítva (naptárhatással nincs kiigazítva), százalékos formában szerepelnek a munkanélküliségi ráta értékei.

A harmonizált fogyasztói árindex (*HICP: Harmonized Index of Consumer Prices*) a fogyasztók által a közönséges áruk kosarában lévő tételekért fizetett végső költségek listája. Így a harmonizált fogyasztói árindex a háztartások által megvásárolt fogyasztási cikkek és szolgáltatások árának időbeli változását méri. Ez pedig a teoretikus modellek között legtöbb esetben filozófiai, szemléletbeli különbség van. mértékét adja meg, mivel harmonizált definíciók szerint számítják ki. A harmonizált fogyasztói árindex jelentőségét mutatja, hogy ezzel az összetett mérőszámmal mérik az inflációt az Európai Unióban.

A harmonizált fogyasztói árindex által mért árak a városi és vidéki árazási minták reprezentatív termékeinek az áraiból származnak. Az index olyan áruk (termékek) árát követi nyomon, mint például a kávé, dohány, hús, gyümölcs, háztartási gépek, autók, gyógyszerek, ruházat, illetve olyan szolgáltatásokat, mint például a víz, gáz, villany, tömegközlekedés, vendéglátás, vagy épp a biztosítási szolgáltatások különböző fajtái. Az index azonban nem tartalmazza például a lakhatási költségeket.

A harmonizált fogyasztói árindex segítségével mért infláció egy adott valuta vásárlóerejének időbeli csökkenését mutatja. Az általános árszint – gyakran százalékban kifejezett – emelkedése azt jelenti, hogy egy valutaegységért ténylegesen kevesebbet lehet vásárolni, mint a korábbi időszakokban. Így az infláció segítségével megismerhetjük egy adott

ország gazdaságának az általános áralakulását, ezáltal pontosabb képet kapva a gazdaság aktuális helyzetéről (erejéről).

Az Európai Központi Bank és az Európai Unió statisztikai hivatala is ezt az összetett mérőszámot alkalmazza az infláció mérésére. Emiatt választottam harmadik látens, gazdasági faktoromnak az inflációt, melyet a harmonizált fogyasztói árindex segítségével mértem.

Az ehhez szükséges adatokat az Eurostat: *HICP - monthly data (index)* nevű adatbázisából (melynek az egyedi online adatkódja: *PRC_HICP_MIDX*) töltöttem le, ahol 2000-től havi bontásban, százalékos formában szerepelnek a harmonizált fogyasztói árindex értékei. A bázisérték a 2015-ös év átlagos havi árindex szintje.

Az elemzés lefolytatásához negyedéves adatokra van szükségem, mivel a fenti két látens változó esetén is negyedéves bontásban voltak elérhetőek az aggregált adatok. Ezért három havonta átlagoltam az infláció mértékét, ezzel pedig megkaptam a negyedévek átlagos inflációs szintjét, melyet már be tudtam építeni az adatbázisomba.

3.3. Az összekapcsolt adatbázis bemutatása

Witzany modellje aggregált adatokra épül és ehhez az Eurostatról letöltött általános gazdasági helyzetet (konjunktúrát) leíró változók negyedéves bontásban szerepelnek, ezért a Bondora adatbázis adatait aggregálnom kellett negyedévenként. Ez azért is tűnt jó döntésnek, mert így minden ország és negyedév esetén legalább 200–300 (egy-egy negyedévek esetén több, mint 2000) adat átlagolásával kaptam meg az aggregált értékeket az egyedi RR és LGD értékekből. Továbbá a valós negyedéves PD értékeket is legalább 2000–3000 (egy-egy negyedévek esetén több, mint 70 000) aktív hitel alapján számoltam ki (az adott negyedévben nemteljesítővé váló hitelek számát elosztottam az adott negyedévben aktív hitelek számával). Ezek alapján kimondható, hogy kellően nagy mintával dolgoztam ahhoz, hogy a kiszámított negyedéves nemteljesítési arányok (valószínűségek), illetve az adott negyedévben nemteljesítő hitelek átlagos vigaszrátája szignifikáns (megfelelő pontosságú) legyen és jól jellemezze az egyes országokhoz tartozó hitelállományt.

Összeillesztettem a Bondorából és az Eurostatról származó adatokat és így kaptam három különböző adatbázist (észt, finn és spanyol), melyek 2016 első negyedévével tartalmazzák negyedéves aggregátumokat 2021 negyedik negyedévéig (24 időszak). Ez közgazdaságtani, statisztikai értelemben elég hosszú adatsor ahhoz, hogy modellt lehessen rá illeszteni és érdemi módon elemezhető legyen.

A három összetett adatbázist már föl lehet használni a Witzany által meghatározott két faktoros modell felírásához (Witzany, 2011). A végső adatbázisok a következő fontos változókat tartalmazzák:

- *Ország*
- *Negyedév*
- *Valós PD*
- *Átlagos RR*
- *Átlagos LGD*
- *Egy főre jutó GDP*
- *Munkanélküliségi ráta*
- *Infláció (harmonizált fogyasztói árindex alapján)*

Az így kapott három adatbázis már minden olyan változót (adatot) tartalmaz, amely elengedhetetlen a modellezéshez. A következő fejezetben bővebben bemutatom ezen adatállományok dinamikáját és érdekesebb jellemzőit. A függelékben pedig megtalálhatók az egyes esetekhez (országokhoz) tartozó összegző statisztikai táblázatok.

3.4. Bővített adatbázis bemutatása (bimodalitás vizsgálata)

A vigaszráta eloszlásának modellezéséhez és ebből következően a bimodalitás (multimodalitás) tulajdonság vizsgálatához összeállított adatbázisom is a Bondora adatbázison alapszik. A korábbiakhoz hasonlóan, ebben az esetben is a 2016 és 2021 közötti időintervallumra szűrtem le a kiindulási adatkészletet. Ennek két praktikus előnye volt. Egyrészt kiküszöbölöm a korábban már említett pontatlanságokat (például a modellkalibrációs időszak hiányos vagy rossz értékeit). Másrészt a két modellezési eljárás – *wrong-way risk* modellek, illetve regressziós (bimodalitást vizsgáló) modellek – összehasonlítását segíti, ha ugyanolyan adatokon és egyforma időszakon alapulnak a mögöttes számítások.

Pontosabban fogalmazva, az adattisztítás során minden olyan hitelszerződést (hitelt) kiszűrtem, amelyek vagy 2016 előtt lezárult (visszafizetésre került vagy nemteljesítővé vált) vagy 2021 után (2022-ben) kötöttek meg. Továbbá kiszűrtem minden olyan tételt, amely esetén a *credit ratingek* alapján becsült PD érték 0 volt. Ezután kiszűrtem a szlovák hiteleket, amely esetén nincs elegendő adat ahhoz, hogy érdemi elemzést lehessen végezni. Végül pedig kiszűrtem minden nem *default-os* hitelt, mert csak a nemteljesítő hitelek esetén rendelkezünk pontos vigaszráta (RR) és nemteljesítés miatti veszteség (LGD) adatokkal.

A regressziós eljárások elvégzéséhez szükség van ún. magyarázó változókra, amelyek a hitelekhez kapcsolódó különböző (szocioökonometriai) jelenségeket írnak le, például a hitelfelvevő életkorát vagy épp a havi törlesztőrészlet értékét. Ezért az alábbi változókat válogattam ki közgazdaságtani és pénzügyi megfontolások szerint a rendelkezésre álló lehetőségek – a Bondora adatbázis 112 változója – közül:

- *ProbabilityOfDefault*
- *EADI*
- *Age*
- *Gender*
- *Education*
- *IncomeTotal*
- *MonthlyPayment*
- *HomeOwnershipType*
- *OccupationArea*

Ahogy már korábban említettem a *ProbabilityOfDefault* (*PD*) változó a nemteljesítés valószínűségét mutatja egy éves időhorizonton belül (*credit-ratings*ek alapján számolva). Az *EADI* változó pedig tartalmazza a nemteljesítéskori kitettség (nemteljesítéskor fennálló tőketartozás) mértékét.

Az *Age* változó megadja a hitelfelvevő (hitelszerződési aláírásakor) életkorát. A *Gender* változó tartalmazza a hitelfelvevő nemét (a következő kódolás alapján: 0 = férfi, 1 = nő és 2 = meghatározatlan). Az *Education* változó mutatja a hitelfelvevő legmagasabb végzettségét. A *IncomeTotal* változó mutatja a hitelfelvevő teljes bevételének nagyságát. A *MonthlyPayment* változó tartalmazza azon becsült összeget, amelyet a hitelfelvevőnek havonta fizetnie kell (törlesztőrészletként). A *HomeOwnershipType* változó megadja, hogy a hitelfelvevő milyen (például saját vagy bérelt) ingatlanban lakik. Végezetül a *OccupationArea* változó mutatja, hogy a hitelfelvevő milyen szektorban dolgozik.

A kiválogatott változók között nemcsak mennyiségi ismérvek szerepelnek, hanem több minőségi ismérv (kategoriális változó) is van. Az utóbbi csoportba tartozó változókat számszerűsíteniem kellett. Például az *Education* változó esetén az egyes képzettségi szintek közt sorrend állítható föl, így ebben az esetben az egyes szinteknek egy-egy sorszámot adok, így egy ordinális (sorrendi) skálán mozgó változót kapok. A *Gender* vagy a *HomeOwnershipType* (kategoriális) változók esetén csak névleges (nominális) mérési skálát lehet alkalmazni, ezért a regresszió eljárások során felhasználható dummy változókat alakítok ki belőlük.

A *Gender* változóból, amely három féle értéket vehet föl, két dummy (bináris) változót készítettem, így egyértelműen át tudtam kódolni numerikus változókká. Az első dummy (*D_woman*) akkor vesz föl egyet, ha a hitelfelvevő nő, egyébként nulla az értéke, míg a második dummy (*D_undef*) akkor vesz föl egyet, ha a hitelfelvevő nem binárisként (nem nő és nem is férfi) definiálja magát, egyébként nullát vesz föl a változó értéke.

A *HomeOwnershipType* változó számos lakhatási formát tartalmaz, melyeket két csoportra soroltam (tulajdonos vagy bérlő). Így az új minőségi változómnak két értéke lehetett, ezért elég volt egy új dummy változót (*D_berlo*) definiálnom ahhoz, hogy regressziós eljárásoknál használható numerikus változót kapjak.

A *OccupationArea* változó több mint 20 féle szektort tartalmaz. Ez túl sok ahhoz, hogy hatékonyan lehessen használni a regressziós eljárásoknál, ezért három csoportra bontottam őket. A hitelfelvevők jelentős részénél nem volt meghatározva, hogy mely szektorban dolgoznak, ezért őket egy csoportba soroltam, és e csoport kapott egy dummy változót (*D_egyeb_munka*). A munkavállalók jelentős része vagy fizikai vagy irodai munkát végez, ezért e két csoportba soroltam a „nem egyéb szektorban” dolgozó hitelfelvevőket. Ezek alapján pedig meghatároztam a *D_irodai* dummy változót, amely értelemszerűen egyet vesz föl, ha a munkavállaló irodai munkát végző szektorban dolgozik és nullát, ha nem. Így a fizikai munkát végző hitelfelvevők esetén mindkét dummy változó nullát fog fölvenni, ezáltal egyértelmű lesz a kódolás.

Az így kapott négy adatbázis már minden olyan változót (adatot) tartalmaz, amely elengedhetetlen a modellezéshez. A következő fejezetben bővebben bemutatom a vizsgálták empirikus (valós) eloszlását és annak érdekesebb jellemzőit. A függelékben pedig megtalálhatók az egyes esetekhez (országokhoz) tartozó összegző statisztikai táblázatok.

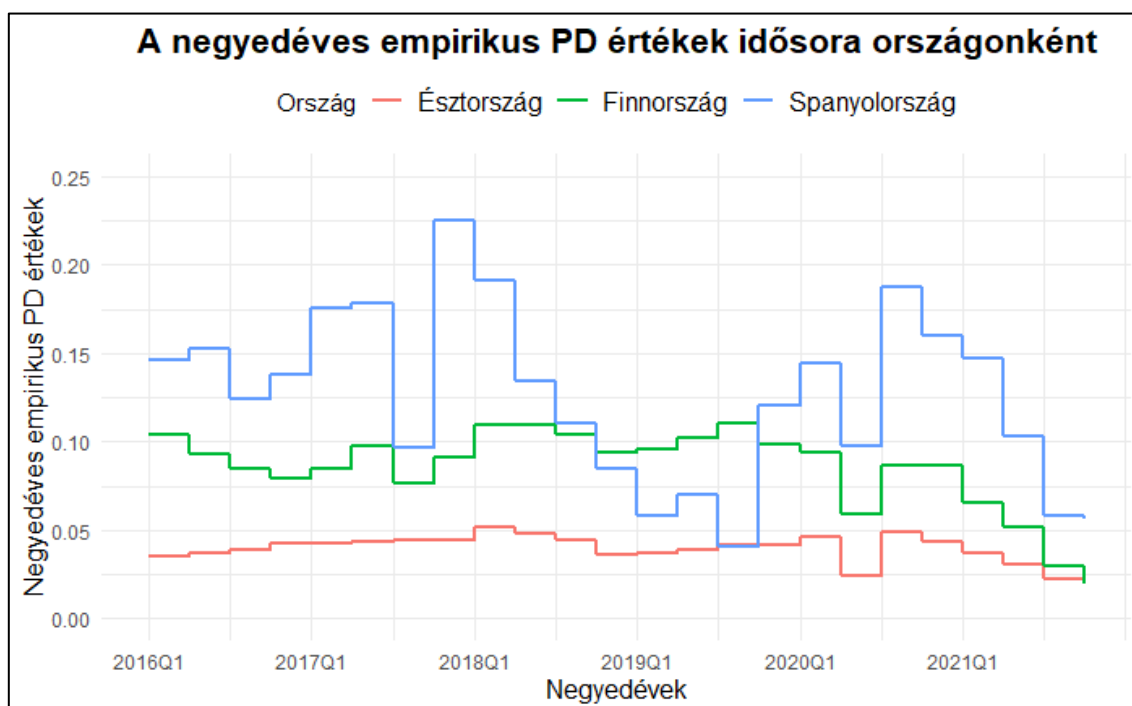
4. Empirikus kutatás és modellezés

Mielőtt a modellezésre térnék, úgy gondolom, hogy érdemes elemezni az előző fejezetben bemutatott adatbázisok főbb változóinak viselkedését. Például hogyan változik időben a vigaszráta átlagos értéke vagy milyen alakú a vigaszráta tapasztalati sűrűségfüggvénye az egyes országok esetén. Ezért a következő alfejezetben részletesebben, ábrákkal illusztrálva tanulmányozom a modellekhez kapcsolódó legfontosabb jelenségeket.

4.1. Adatbázisok elemzése

A *wrong-way risk* modellezéshez kapcsolódó adatbázis két legjelentősebb eleme a negyedévenkénti nemteljesítések aránya – amit az adott negyedévben történő nemteljesítés valószínűségként (PD) is értelmezhetünk – és az adott negyedévben nemteljesítő hitelek vigaszrátájának negyedévenkénti átlagos értéke. Az alábbi ábrán a három vizsgált ország esetén látható az empirikus PD értékek időszora (azért nevezem empirikus PD-nek, mert ezen értékek az adott negyedévben nemteljesítővé váló hitelek arányát mutatja, nem pedig *credit-ratingek* alapján becsült értékeket illusztrálja):

1. ábra: A negyedéves empirikus PD változása időben az egyes országok esetén



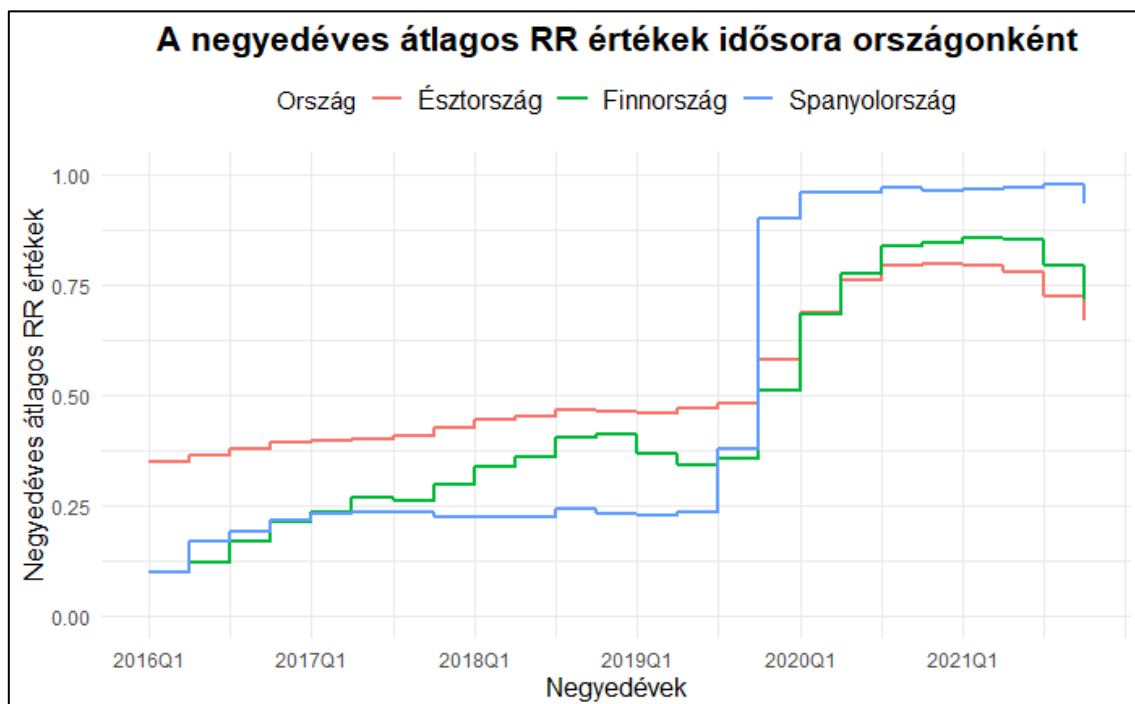
Az ábrán jól látható mennyire különbözik a három idősor dinamikája. A vizsgált három ország közül Észtországban a legalacsonyabb az empirikus PD értéke. A megfigyelt 24 negyedév során kis eltéréssel mindig 4% körüli értékeket fejt föl. Az utóbbi 4-5 negyedévben egy enyhe javulás – PD csökkenése – látható, de nagy vonalakban a dinamikára nézve kijelenthető, hogy majdnem konstans az empirikus PD idősora.

Ezzel szemben Finnország empirikus PD idősora sokkal érdekesebb képet fest. Ebben az esetben egy javuló tendenciát – csökkenő PD – figyelhetünk meg, amely körül szemmel láthatóan ingadoznak az értékek. Az idősor 10%-ról indul és sokáig e körül – az észt adatokhoz képest jelentősen – ingadoznak az értékek, majd 2019 harmadik negyedétől kezdődően nagy mértékű csökkenés fedezhető fel.

A három idősor közül a spanyol a legérdekesebb, mert ebben az esetben rettentő nagy, akár 10% pontos változásokat is megfigyelhetünk az egyes negyedévek között. Ezáltal egy nagyon szokatlan dinamikát láthatunk az ábrán, amely mögött feltételezhetően több, nem igazán mérhető gazdasági és szociális folyamat húzódhat meg.

Az 1. ábra segítségével bemutattam az egyes országok esetén hogyan alakulnak az empirikus PD-k idősorai, most pedig áttérek a *wrong-way risk* modellekhez kapcsolódó másik érdekes jelenségre, a vigaszráták időbeni alakulására. Felmerül a kérdés, hogy esetükben az elméletnek megfelelően negatívan korrelálnak-e PD és RR értékekkel, azaz, ha a nemteljesítés valószínűsége nő, akkor a vigaszráták átlagos értéke csökken-e?

2. ábra: A negyedéves átlagos RR változása időben az egyes országok esetén



Ha összehasonlítjuk az 1. és 2. ábrát, akkor ez a kapcsolat nem olvasható le egyértelműen. Azonban számos érdekesség látható az utóbbi ábrán. Például a 2020 elején látható nagyobb ugrástól eltekintve mindhárom ország esetén egy stabil és lassú javulás látható (főként 2016 és 2019 között) a megfigyelt időszak alatti PD ingadozással szemben. További érdekesség, hogy a koronavírus járvány felfutása ellenére a nemteljesítés valószínűsége csökkent és a vigaszráták átlagos értéke jelentősen nőtt, mindhárom vizsgált országban.

Az adatbázis behatóbb vizsgálata utána arra a konklúzióra lehet jutni, hogy a spanyol adatoknál 2019 harmadik negyedében bekövetkezett nagyjából öt tizednyi ugrás annak köszönhető, hogy a Bondora 2019 elején feltehetően egy új hitelterméket vezetett be a piacra, amely teljes fedezettséggel rendelkezik. Ezt a következtetést az a tény támasztja alá, hogy a 2019 végétől nemteljesítővé váló, 1-es vigaszrátával rendelkező hitelek nagy többségét 2019-től kezdődően kötötték. Továbbá a *default-ossá* váló hitelek között megnő az 1-es vigaszrátával rendelkezők aránya.

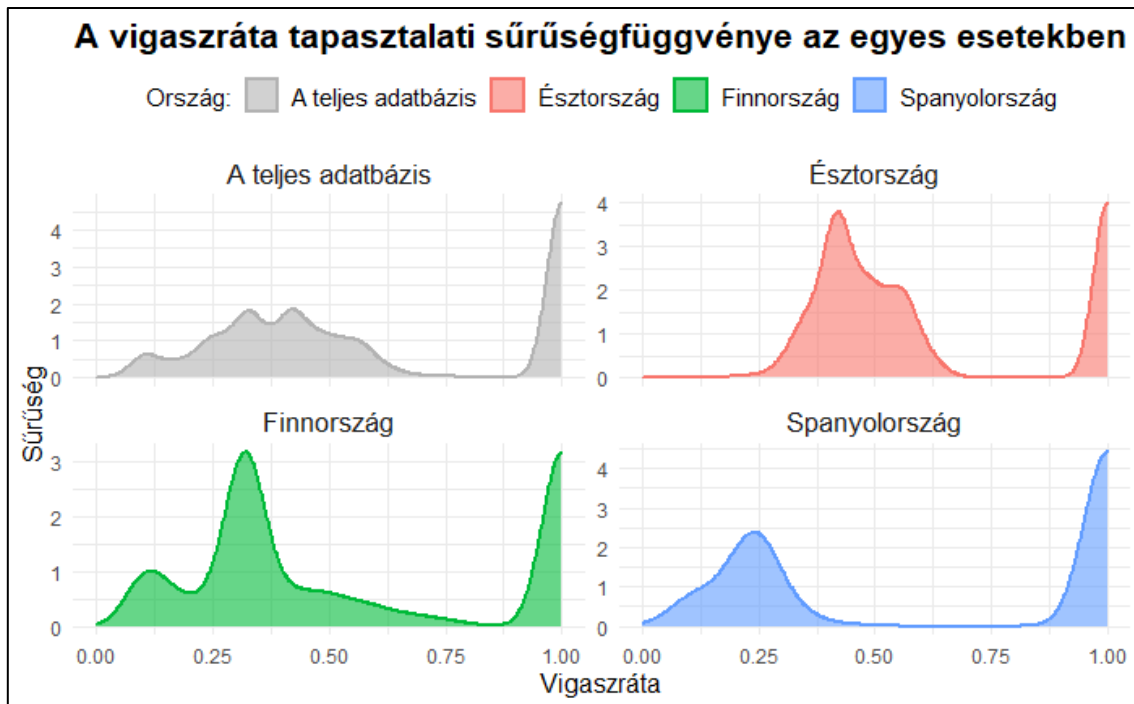
Észtország és Finnország adatait vizsgálva is észrevehető egy trendhez képest nagyobb ugrás 2019 végén, 2020 elején, azonban nem olyan számottevő, mint Spanyolország esetén. Itt a vigaszráták átlagos értékében két-három negyedév alatt realizálódik a jelentősebb változás. Észtország és Finnország esetén is a jelenség mögött álló magyarázat feltehetően az előbb említett termék megjelenése és a koronavírus járvány által okozott gazdasági válság miatti nehézségek.

Az 1-es és 2-es ábrák alapján egy nagyon érdekes időszakot vizsgállok, ahol számos figyelemreméltó esemény következett be, ezért kíváncsi vagyok, hogy mit mutat majd az empiria. Ezen adatok esetén is fennállnak a korábban bemutatott elméleti összefüggések, illetve, hogy miként tudja ezt az adatsort kezelni Witzany modellje.

Mielőtt a modellezésre térnék, úgy gondolom érdemes áttekinteni a teljes adatbázis, illetve az egyes országok esetén külön-külön a vigaszráták eloszlását. Ennek megjelenítésére a legjobb grafikon a kernel alapú sűrűségfüggvényt használó diagram, amely egyetlen folytonos görbe segítségével mutatja a mögöttes adatsor sűrűségfüggvényét, ezáltal egyértelműen meghatározva annak eloszlását.

A kernel típusú (simított) sűrűségfüggvény diagramja hasonló a hisztogramhoz, de finomabban ábrázolja az eloszlás alakját, mivel a hisztogram esetén használt csoportosítás (*binnig*) szélessége nincs rá hatással. Az alábbi ábrán jól látható a teljes adatbázison, illetve csak az észt, a finn és a spanyol adatokon alapuló vigaszráták kernel típusú (simított) tapasztalati sűrűségfüggvénye:

3. ábra: A vigaszára tapasztalati sűrűségfüggvénye az egyes esetekben (országokén)



Az ábrázolás során próbálok konzisztenciára törekedni. Ezért mindent, amit a teljes adatbázis alapján illusztrálok az szürke színű. Ugyanígy, amit Észtország adatai alapján szemléltetek az mindig piros színű, amit Finnország adatai alapján az mindig zöld, és amit Spanyolország adatai alapján az mindig kék.

Az ábrán jól látható, hogy egészen különböző alakú sűrűségfüggvényeket kaptunk az egyes esetekben. Két fontos hasonlóságot emelnék ki. Egyrészt, mind a négy esetben a 0 és 1 közötti intervallumon vesznek föl értékeket az empirikus sűrűségfüggvények. Másrészt, az összes esetben jelentős koncentráció látható az $x = 1$ határpont környékén. Ezekről eltekintve igen különböző sűrűségfüggvényeket láthatunk.

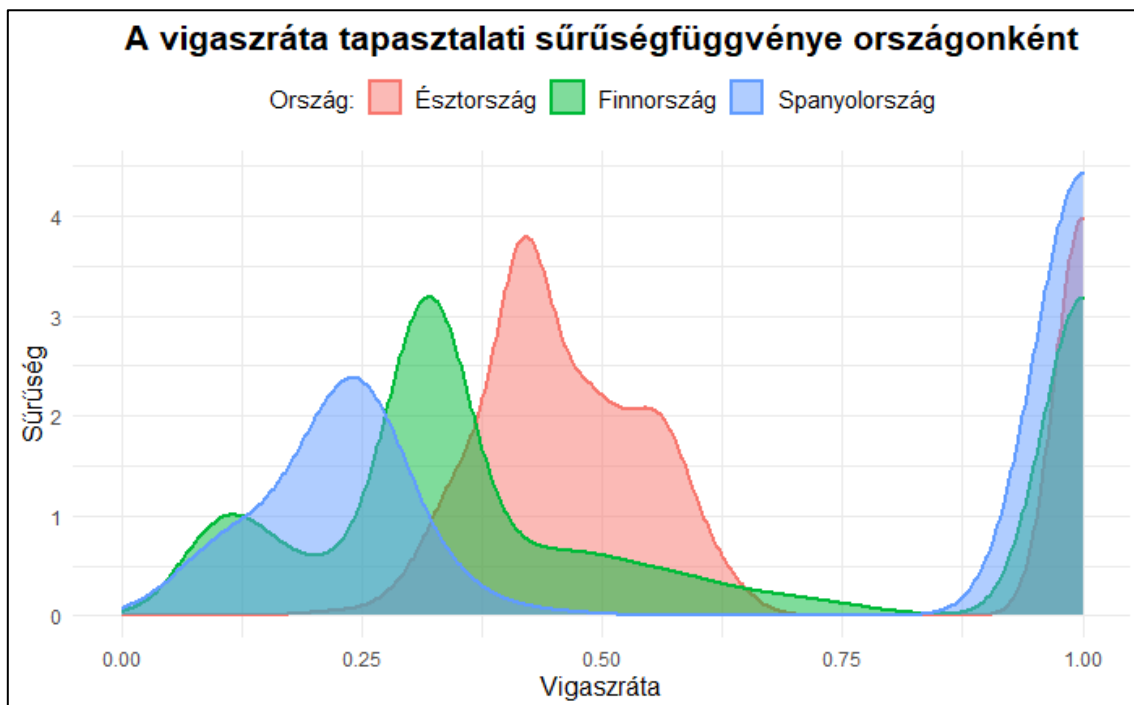
Spanyolország esetén egyértelműen kirajzolódik a bimodalitás, amiről a bevezetőben már szót ejtettem. Ebben a vonatkozásban nagyjából egyezik az elmélet az empiriával. Észtország esetén is megközelítőleg teljesül a bimodalitás, bár 0,6 környékén látható egy kis kiszélesedés. Finnország esetén azonban már láthatóan sérül a bimodális tulajdonság és nagyjából 0,10 és 0,15 között megjelent egy harmadik „púp”, így ebben az esetben már a tapasztalati sűrűségfüggvény trimodális, amiről szintén volt szó a bevezetésben, illetve a módszertani fejezetben.

Az elmélettől leginkább különböző eset az, amikor a teljes adatbázis vigaszárait vizsgálom. Az $x = 1$ határpont mellett 0,25 és 0,50 között látható két kisebb „púp” is, ami alapján multimodális tulajdonságúnak lehet nevezni ezt a tapasztalati sűrűségfüggvényt.

Azonban, ha erősebben simított kernel görbét alkalmazok, akkor ez az eset is bimodálissá válik. Ha nagy vonalakban ránézünk, akkor már most is kirajzolódik egy „nagyobb púp”, hozzávetőleg 0,20 és 0,75 között.

Érdemes egy ábrán megjeleníteni és összehasonlítani a három ország tapasztalati sűrűségfüggvényeit. Ekkor láthatóvá válik számos fontos hasonlóság és különbség. Például az $x = 1$ határpont környékén egész szépen fedik egymást a kernel típusú sűrűségfüggvények. Azonban a másik sűrűsödési pont igen különböző helyen látható az egyes országos estén. Vegyük észre, hogy mindhárom esetben alig van olyan nemteljesítő hitel, amelynek a vigaszártája 0,75 és 0,85 között van.

4. ábra: A vigaszártá tapasztalati sűrűségfüggvénye országonként (együttesen ábrázolva)



Az ábrát nézve két kérdés jut eszembe:

- A. Hogyan hat az egyes országok esetén a vigaszártá eloszlása a korrelációs paraméterekre?
- B. Az ennyire U alakra hajazó görbét meg tudják-e ragadni a különböző regressziós módszerek?

Az első kérdésre a következő 4.2-es alfejezetben a Q függvény – vigaszártá hosszútávú eloszlásfüggvényének – vizsgálata közben térek ki. Míg a második kérdésre a 4.3-as alfejezetben keresek választ és ott bővebben kifejtem a modellezésem eredményeit és tapasztalatait.

4.2. Wrong-way risk modellezés (korreláció vizsgálata)

A módszertani fejezetben bemutatott *wrong-way risk* modell fő célja, hogy a kulcsfontosságú paramétereket (ρ_1 , ρ_2 és ω) megbecsülje a fent bemutatott elmélet és a megfigyelt PD és RR vagy LGD idősorok alapján. Az elemzésem során a vigaszráta adatokat fogom használni (Witzany a nemteljesítés miatti veszteségekre alkalmazta a modelljét), azonban mindkét jelenségre ugyanúgy működik az eljárás, és várhatóan csak az *omega* paraméter előjele változik az elméletnek megfelelően.

Ahhoz, hogy a modellt tudjuk illeszteni a megfigyelt adatsorokra szükséges meghatározni az RR hosszútávú eloszlásfüggvényét (*Q függvény*) és a nemteljesítés (PD_0) szintjének y_D transzformáltját. Továbbá meg kell határozni a két *ARMA*(p, q) folyamat együtthatóit. Ezek alapján meghatározható a *likelihood* függvény:

- $\mathcal{L}(\langle pd(t); lgd(t) \rangle \mid \rho_1; \rho_2; \omega; G; y_D; \alpha_{k,i}; \beta_{k,j})$

Ez egy nagyon komplikált két változós becslőfüggvény, amit nagyon nehéz egyben kifejezni, ezért a cseh kutató levezette a következő két *likelihood* függvényt:

- $\mathcal{L}(pd(t) \mid pd(\tau), \tau < t)$
- $\mathcal{L}(lgd(t) \mid pd(\tau), \tau \leq t; lgd(\tau), \tau < t)$

melyek segítségével felbontható a fent definiált *likelihood* függvény. Pontosabban felbontható e két (egyváltozós) *likelihood* függvény produktumára a felső bonyolult összefüggés (Witzany, 2009):

$$(28) \quad \mathcal{L}(\langle pd(t); lgd(t) \rangle \mid \rho_1; \rho_2; \omega; G; y_D; \alpha_{k,i}; \beta_{k,j}) = \\ = \prod_{t=1}^T \mathcal{L}(pd(t) \mid pd(\tau), \tau < t) \cdot \mathcal{L}(lgd(t) \mid pd(\tau), \tau \leq t; lgd(\tau), \tau < t)$$

A kezelhetőség és rövidség kedvéért ezentúl nem írom ki az összes feltételes paramétert. Witzany elméleti modellje alapján pedig definiálható a két *likelihood* függvény az alábbiak szerint:

$$(29) \quad \mathcal{L}(pd(t) \mid pd(\tau), \tau < t) = \frac{\varphi(u_1(t))}{\beta_{1,0} g'(x_1(t))}$$

$$(30) \quad \mathcal{L}(lgd(t) \mid pd(\tau), \tau \leq t; lgd(\tau), \tau < t) = \frac{\varphi(u_2(t))}{\sqrt{1-\omega^2} \beta_{2,0} h'(\omega x_1(t) + \sqrt{1-\omega^2} x_2(t))}$$

Így a (28) egyenlet átalakítható a modellezésnél használható formába:

$$(31) \quad \mathcal{L} = \prod_{t=1}^T \frac{\varphi(u_1(t))}{\beta_{1,0} g'(x_1(t))} \prod_{t=1}^T \frac{\varphi(u_2(t))}{\sqrt{1-\omega^2} \beta_{2,0} h'(\omega x_1(t) + \sqrt{1-\omega^2} x_2(t))}$$

Először a két függvény deriváltját vezetem le paraméteresen, majd ezek után számolom ki a konkrét behelyettesítendő (x_1 és x_2 , u_1 és u_2 , $\beta_{1,0}$ és $\beta_{2,0}$) értékeket. A módszertani részben definiált g függvény – (7) egyenlet – deriváltja a láncszabály alkalmazásával könnyen kiszámolható.

$$(32) \quad \frac{\partial}{\partial x} g(x) = \sqrt{\frac{\rho_1}{1-\rho_1}} \varphi\left(\frac{\sqrt{\rho_1}x - y_D}{\sqrt{1-\rho_1}}\right) = g'(x)$$

Ennél viszont sokkal nehezebb feladat a (8) egyenletben szereplő h függvény deriváltjának a kiszámítása. A függvény argumentumában szereplő G függvényt az alábbiak szerint határozta meg Witzany:

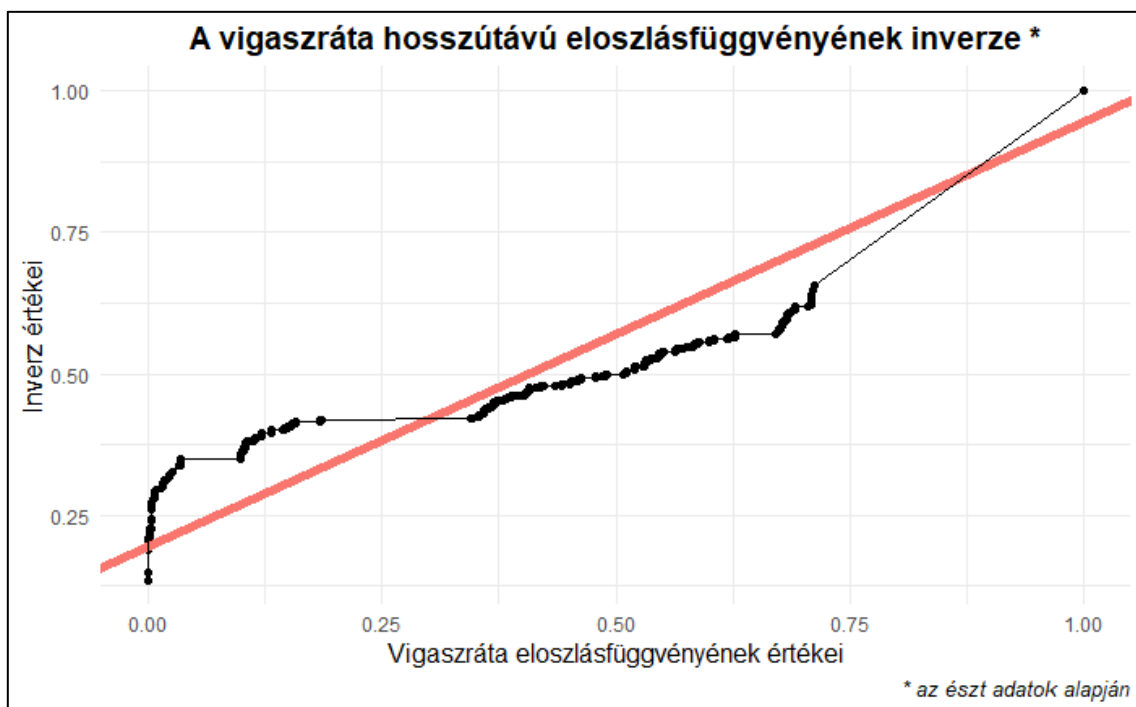
$$(33) \quad G = Q^{-1} \circ \Phi \Rightarrow G(x) = Q^{-1}(\Phi(x))$$

ahol Q a vigaszráta hosszútávú eloszlásfüggvényét, Φ pedig a standard normális eloszlás eloszlásfüggvényét jelöli. Q inverze nagyjából lineáris (kvázilineáris) a $[0, 1]$ intervallumon mindhárom ország esetén (Witzany ezt a megfigyelést használta fel a feladat egyszerűsítésére), ezért az inverzfüggvény közelíthető az alábbiak szerint:

$$(34) \quad Q^{-1}(x) \approx m \cdot x + b$$

A közelítést igen jól mutatja a következő (5. ábra), ahol az észti RR adatok alapján ábrázoltam a vigaszráták eloszlásfüggvényének inverzét. Pirossal jelöltem a függvénygörbét közelítő egyenest, ami úgy viselkedik, mint egy görbéhez illesztett lineáris trendvonal.

5. ábra: A vigaszráta hosszútávú eloszlásfüggvényének inverzfüggvénye



A piros egyenes egyenlete a következő, amelyet fel fogok használni a (34) egyenlet alapján a továbbiakban a h függvény deriváltjának kiszámításához:

$$(35) \quad Q^{-1}(x) \approx 0.7508234 \cdot x + 0.1948296$$

A három ország esetén külön-külön is meghatározhatnám a Q inverzét, azonban az egyszerűség kedvéért mindhárom esetben ezt az egyenletet fogom használni, hogy a modellek könnyebben összehasonlíthatóak legyenek.

Ezek alapján a (8) egyenletben szereplő h függvényt tovább tudom transzformálni, annak érdekében, hogy könnyebben ki tudjam számolni annak x szerinti deriváltját.

$$(36) \quad \begin{aligned} h(x) &= \int_{-\infty}^{\infty} G(\sqrt{\rho_2}x + \sqrt{1-\rho_2}w) \varphi(w) dw \approx \\ &\approx \int_{-\infty}^{\infty} (m \cdot \Phi(\sqrt{\rho_2}x + \sqrt{1-\rho_2}w) + b) \varphi(w) dw = \\ &= \int_{-\infty}^{\infty} (m \cdot \Phi(\sqrt{\rho_2}x + \sqrt{1-\rho_2}w) \varphi(w) + b \varphi(w)) dw \end{aligned}$$

ahol φ a standard normális eloszlás sűrűségfüggvényét jelöli. A most kapott kifejezést már tudom deriválni a paraméteres integrálok „bederiválhatóságáról” szóló tétel alapján. Ezen kívül még azt is felhasználtam, hogy a sűrűségfüggvény teljes értelmezési tartományon vett integrálja egyenlő eggyel.

$$(37) \quad \begin{aligned} \frac{\partial}{\partial x} h(x) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial x} [m \cdot \Phi(\sqrt{\rho_2}x + \sqrt{1-\rho_2}w) \varphi(w) + b \varphi(w)] dw = \\ &= \int_{-\infty}^{\infty} \sqrt{\rho_2} \cdot m \cdot \varphi(\sqrt{\rho_2}x + \sqrt{1-\rho_2}w) \varphi(w) dw = \\ &= \sqrt{\rho_2} \cdot m \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{(\sqrt{\rho_2}x + \sqrt{1-\rho_2}w)^2 + w^2}{2}} dw = \\ &= \sqrt{\rho_2} \cdot m \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{\rho_2 x^2 + 2\sqrt{\rho_2(1-\rho_2)}xw + (1-\rho_2)x^2 + w^2}{2}} dw = \\ &= \sqrt{\rho_2} \cdot m \cdot \frac{1}{2\pi} e^{-\frac{\rho_2 x^2}{2}} e^{\frac{\rho_2(1-\rho_2)x^2}{2(2-\rho_2)}} \int_{-\infty}^{\infty} e^{-\frac{(2-\rho_2)w^2 + 2\sqrt{\rho_2(1-\rho_2)}xw + \frac{\rho_2(1-\rho_2)}{2-\rho_2}x^2}{2}} dw = \\ &= \sqrt{\rho_2} \cdot m \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\rho_2 x^2}{2}} \cdot e^{\frac{\rho_2(1-\rho_2)x^2}{2(2-\rho_2)}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\sqrt{2-\rho_2}w + \sqrt{\frac{\rho_2(1-\rho_2)}{2-\rho_2}}x\right)^2}{2}} dw = \\ &= \sqrt{\rho_2} \cdot m \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\rho_2 x^2}{2(2-\rho_2)}} = h'(x) \end{aligned}$$

Az utóbbi két oldalon kiszámolt deriváltakat behelyettesítem a (31) kifejezésbe. Ekkor egy olyan R-ben is programozható *likelihood* függvényt kapunk, melyet fel tudok használni a modellezés során:

$$(38) \quad \mathcal{L} = \prod_{t=1}^T \frac{\varphi(u_1(t))}{\beta_{1,0} \cdot \sqrt{\frac{\rho_1}{1-\rho_1}} \varphi\left(\frac{\sqrt{\rho_1}x_1(t)-y_D}{\sqrt{1-\rho_1}}\right)} \prod_{t=1}^T \frac{\varphi(u_2(t))}{\sqrt{1-\omega^2} \cdot \beta_{2,0} \cdot \sqrt{\rho_2} \cdot m \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{\rho_2 \cdot (\omega x_1(t) + \sqrt{1-\omega^2}x_2(t))^2}{2(2-\rho_2)}}$$

Most már csak fel kell tölteni a *likelihood* függvényt a megfelelő változókkal. Tehát meg kell határozni az $x_1(t)$ és $x_2(t)$ faktorokat, az $u_1(t)$ és $u_2(t)$ reziduumokat, a $\beta_{1,0}$ és $\beta_{2,0}$ együtthatókat – az utóbbi négyet az $ARMA(p, q)$ modellek alapján – és y_D , amely a hosszútávú nemteljesítési valószínűségnek a transzponáltja. Az utóbbi kifejezésben még szerepel három korrelációs paraméter ω , illetve ρ_1 és ρ_2 , ezeket akarjuk majd megbecsülni *MCMC* szimuláció segítségével.

Először a két faktort (x_1 -t és x_2 -t) határozom meg a három látens változómból főkomponens elemzés (*PCA: principal component analysis*) segítségével. Az így kialakított első főkomponenst a PD szisztematikus faktoraként (x_1), míg a második főkomponenst az RR szisztematikus faktoraként értelmezem (x_2). Ez egy eléggé leegyszerűsített megoldás a két faktor kialakítására, azonban a fő célom az, hogy az eljárást (a módszertant) bemutassam, nem pedig, hogy a lehető legpontosabban faktorokat alakítsak ki és azok segítségével a lehető legpontosabb becslést adjak a korrelációs paraméterekre. A módszer több látens faktor segítségével is hasonlóan működne (x_1 -nek és x_2 -nek akár külön is lehet gyűjteni látens faktorokat a pontosság növelése érdekében).

Az elmélet szerint $x_1(t)$ és $x_2(t)$ standard normális eloszlású valószínűségi változók, ezért a látens változóimat (egy főre jutó bruttó hazai termék, munkanélküliségi ráta, infláció) standardizáltam és a főkomponens elemzés segítségével kialakítottam a két új faktort a következők főkomponenssúlyok szerint (szorzatösszegként):

4. táblázat: Főkomponenssúlyok táblázata (észti adatok)

	<i>Egy főre jutó GDP</i>	<i>Munkanélküliség</i>	<i>GDP</i>
PD faktora (x_1)	0,9595860	-0,462973325	0,9674378
RR faktora (x_2)	-0,2314237	-0,886350008	-0,1946228

5. táblázat: Főkomponenssúlyok táblázata (finn adatok)

	<i>Egy főre jutó GDP</i>	<i>Munkanélküliség</i>	<i>GDP</i>
PD faktora (x_1)	0,9273472	-0,80623235	0,9347771
RR faktora (x_2)	-0,2738360	-0,59147898	-0,2384830

6. táblázat: Főkomponenssúlyok táblázata (spanyol adatok)

	<i>Egy főre jutó GDP</i>	<i>Munkanélküliség</i>	<i>GDP</i>
PD faktora (x_1)	0,76835831	-0,9207281	0,9167535
RR faktora (x_2)	0,63998699	0,2585506	-0,2767206

A három táblázatot összehasonlítva észrevehetjük, hogy a PD faktora (x_1) mindhárom esetben hasonló szerketű, bár a munkanélküliség főkomponenssúly erőteljesen változik az egyes esetekben. Ezzel szemben az RR faktorának (x_2) kialakításakor nagyobb különbségeket figyelhetünk meg. Például Észtország és Finnország esetén az egy főre jutó GDP főkomponenssúlya negatív és 25% körüli, addig Spanyolország esetén pozitív és 64% körüli. Ezek alapján az lehet mondani, hogy az észt és a finn faktorok hasonló szerkezetűek, míg ezektől a spanyol adatok faktorai különböznek, főként az x_2 faktor. Kíváncsi vagyok, hogy ez hogyan hat a korrelációs modellezés során és hogy a faktorok szerkezetbeli különbségének lesz-e érezhető hatása.

A látens változókból készült faktoraim esetén is felléphet autokorreláció, ezért ennek a kezelésére a (9) egyenletnek megfelelően $ARMA(p, q)$ modellt illeszttek az $x_1(t)$ és $x_2(t)$ idősorokra. Korrelogramok és az auto.arima függvény segítségével meghatároztam az összes változó esetén a modellspecifikációkat.

7. táblázat: ARMA modellspecifikáció az egyes változók esetén (országoként)

	<i>Specifikáció</i>
Észt PD faktor	ARMA(1, 0)
Észt RR faktor	ARMA(1, 0)
Finn PD faktor	ARMA(2, 1)
Finn RR faktor	ARMA(1, 0)
Spanyol PD faktor	ARMA(2, 1)
Spanyol RR faktor	ARMA(0, 0)

A faktorokra illesztett ARMA modellek segítségével meg tudom határozni a hibatagokat. Az $x_1(t)$ faktor idősorának reziduuma alkotják az $u_1(t)$ idősort, míg az $x_2(t)$ faktor idősorának reziduuma az $u_2(t)$ idősort. Fontos, hogy a hibatagok függetlenek legyenek egymástól. Ezt a Ljung-Box teszt segítségével ellenőrizhetem le. A teszt p -értékei 6,25% és 68,25% között változnak, így szerencsére minden esetben el tudom fogadni (nem vetem el 5%-on) a nullhipotézist, azaz a hibatagok között nem lép föl autokorreláció, tehát a reziduumok függetlenek egymástól.

Eddig sikerült meghatároznom az $x_1(t)$ és $x_2(t)$ faktorokat, illetve az $u_1(t)$ és $u_2(t)$ reziduumokat, most pedig Witzany modelljének legzavarosabb része következik, a $\beta_{1,0}$ és $\beta_{2,0}$ együtthatók meghatározása. Az én értelmezésem szerint Witzany a $\beta_{1,0}$ és $\beta_{2,0}$ együtthatókat az ARMA modell béta paramétereinek alapján határozta meg az alábbiak szerint:

$$(39) \beta_{k,0} = \begin{cases} \sqrt{1 - \beta_{k,max}^2} & \text{ha van MA tagja az illesztett modellnek} \\ 1 & \text{ha nincs MA taggal az illesztett modellnek} \end{cases}$$

ahol $\beta_{k,max}$ a legnagyobb késleltetésű $MA(q)$ tag együtthatója (q max).

Már csak az y_D kiszámítása maradt hátra. Ehhez vennem kell az egyes adatsorok esetén, hogy hány hitelszerződés vált nemteljesítővé és elosztanom a vizsgált időszakban aktív hitelek számával, majd az alábbi képletnek megfelelően transzformálnom:

$$(40) y_D = -\Phi^{-1}(PD_0)$$

Ennek a megjelenítésére készítettem egy informatív táblázatot:

8. táblázat: Nemteljesítés valószínűségének (arányának) hosszútávú értéke az egyes esetekben (országok)

	<i>nemteljesítő hitelek száma</i>	<i>aktív hitelek száma</i>	PD_0	y_D
Észtország	31 235	119 137	0,2621	0,6366
Finnország	21 092	56 089	0,3760	0,3158
Spanyolország	17 742	25 797	0,6877	-0,4894

A táblázatból számos meglepő dolgot olvashatunk le. Spanyolországban a vizsgált időszakban a Bondora hiteleinek több, mint 68%-a vált nemteljesítővé, ami egy hatalmas arány. A legkisebb PD_0 -al rendelkező ország, Észtország esetén is nagyon magas a (26%) a default-os hitelek aránya. Érdekes kérdés, hogy ezek a nagyon magas nemteljesítési arányok hogyan

befolyásolják a PD és az RR faktorai között fennálló korrelációt? (A Witzany által vizsgált adatsor esetén PD_0 értéke mindösszesen 0,36% volt!)

Most már megvan minden változó és minden függvény a *likelihood* függvény felírásához és a becslés elvégzéséhez. A ρ_1 , ρ_2 és ω paramétereket a *bayes-i MCMC (Markov Chain Monte Carlo)* eljárás segítségével közelíthetjük.

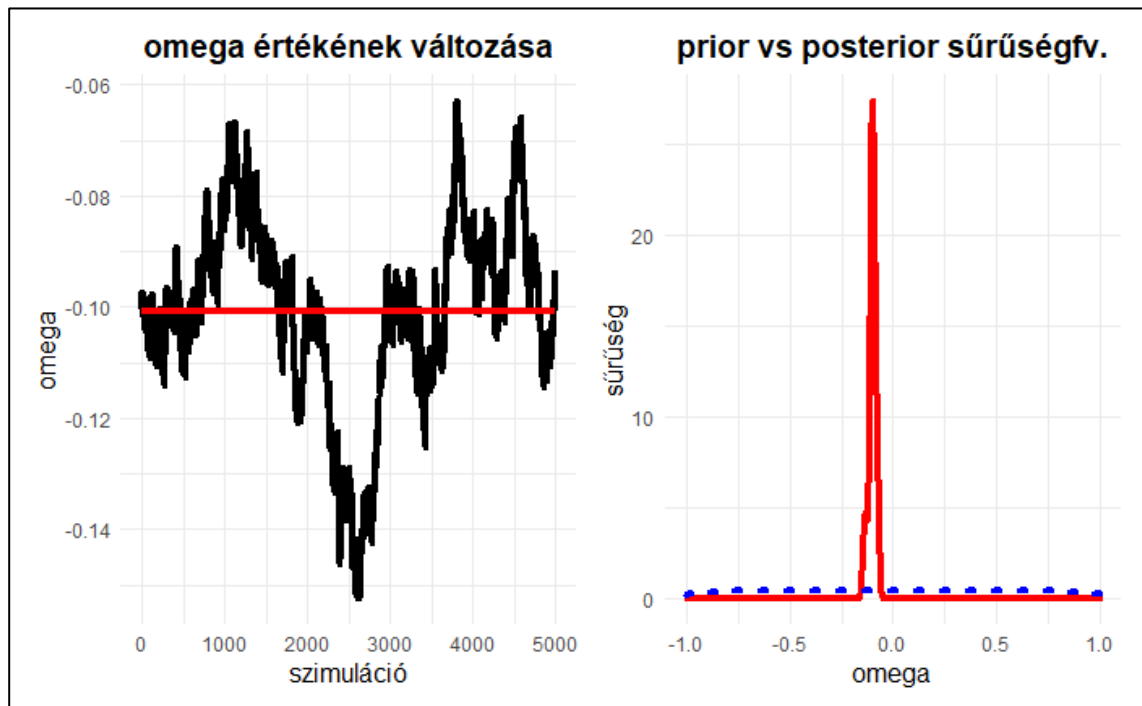
Általánosságban elmondható a *bayes-i MCMC* mintavételi algoritmus családról, hogy ez egy nagyon elterjedt és gyakran használt eszköze a többdimenziós paraméterekkel rendelkező összetett modellek becslésének. Ilyen modellek közé tartoznak az ugró folyamatok, sztochasztikus volatilitás modellek vagy az összetett korrelációs szerkezetű pénzügyi sztochasztikus modellek. A legtöbb *MCMC* eljárás szekvenciális mintavételre alapul, amely alapján a vizsgált paraméter (vagy paraméterek) *posterior* eloszlását közelíti és a közelítés segítségével korrigálja a későbbi mintavételeket. A módszer legkritikusabb pontja a paraméter konvergenciája, minden szimuláció esetén felmerül az a kérdés, hogy a paraméter ténylegesen közelíti-e a valós „céeloszlást”? A *MCMC* típusú módszereket mutatta be részletesen az alábbi két publikáció (Johannes & Polson, 2010; Lynch, 2007).

A paraméterbecsléshez a *MCMC* módszerek családjába tartozó *Gibbs sampler* eljárást használtam, melyet Witzanyhoz hasonlóan összekapcsoltam a *Metropolis–Hastings* algoritmussal. A paraméterekre vonatkozó feltételes valószínűségeket úgy kapjuk meg, hogy a Bayes-tételt alkalmazzuk a *likelihood* függvényre és egy *posterior* sűrűségfüggvényre. Egy uninformatív *prior-t* – $\beta(1, 1)$ eloszlást – választottam, ami lényegében az egyenletes eloszlás a $[0, 1]$ intervallumon, ezáltal a *prior-nak* nincs hatása a *posterior* eloszlásra.

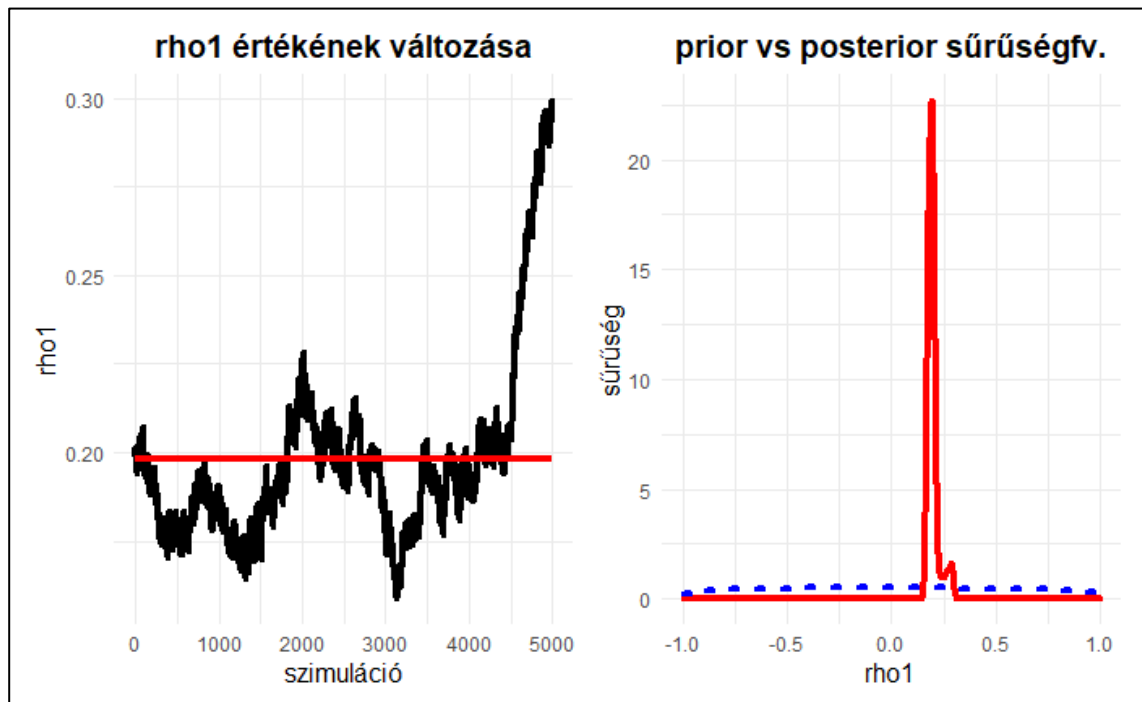
Az eljárás lényege az, hogy egy iterációs körben először az első paraméterre adok egy becslést úgy, hogy az adott paramétert egy kicsi értékkel megváltoztatom. Ekkor meg tudjuk vizsgálni, hogy a paraméter megváltoztatása növelte-e a paraméter *likelihood*-ját; ha igen akkor a megváltoztatott paraméter értéke kerül be az új paraméter értéknek, ha nem javult, akkor marad a régi. Ezt az algoritmus elvégzem egyesével minden vizsgált paraméterre egy iterációs körön belül, majd ezt a kört ismételve addig, míg a paraméterek korrelációja nem lesz megfelelő.

A *maximum likelihood* becslésnek számos hátránya van. Például nagyon érzékeny a kezdeti értékre, nagyon sok esetben a kiindulási értékhez konvergál az érték, ami a kiindulási érték változtatásával problémássá válik. Továbbá érzékeny az eljárás során használt „érzékenységi” paraméterre (lépésközre), amely alapján az iterációnál egy lépésben megváltoztatok egy paraméter értéket. E két tulajdonságot igazán jól szemléltette az általam futatott paraméterbecslés.

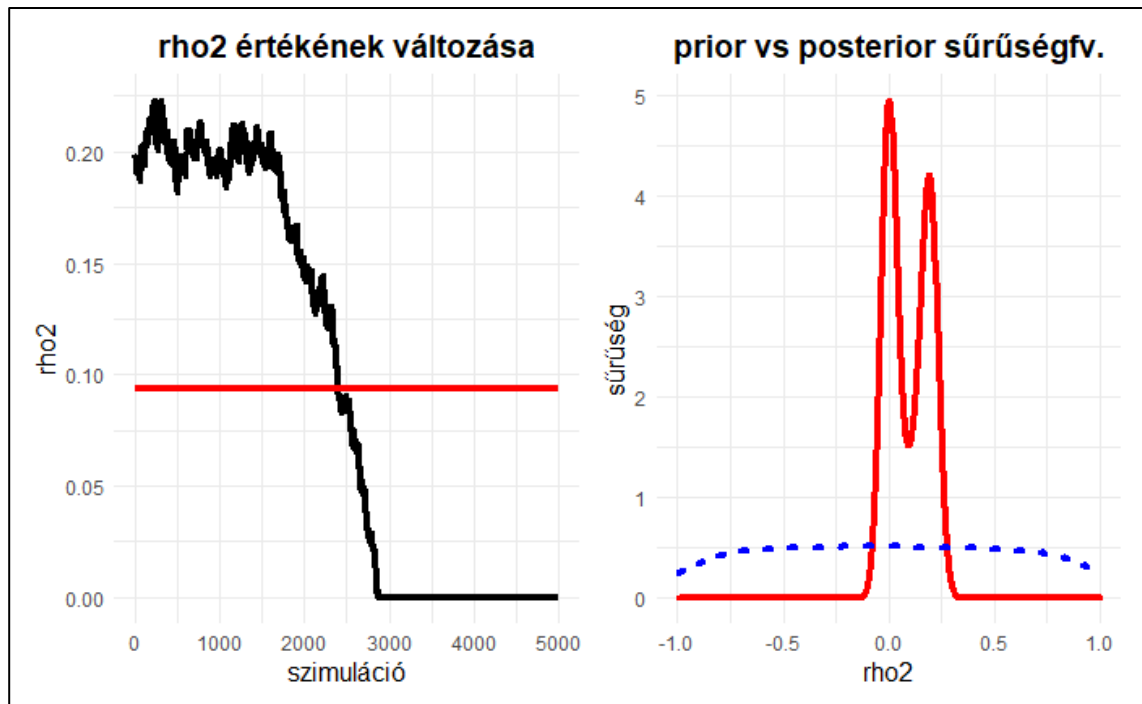
6. ábra: Az omega paraméter konvergenciája és eloszlása (az észl adatsor alapján)



7. ábra: A rho1 paraméter konvergenciája és eloszlása (az észl adatsor alapján)



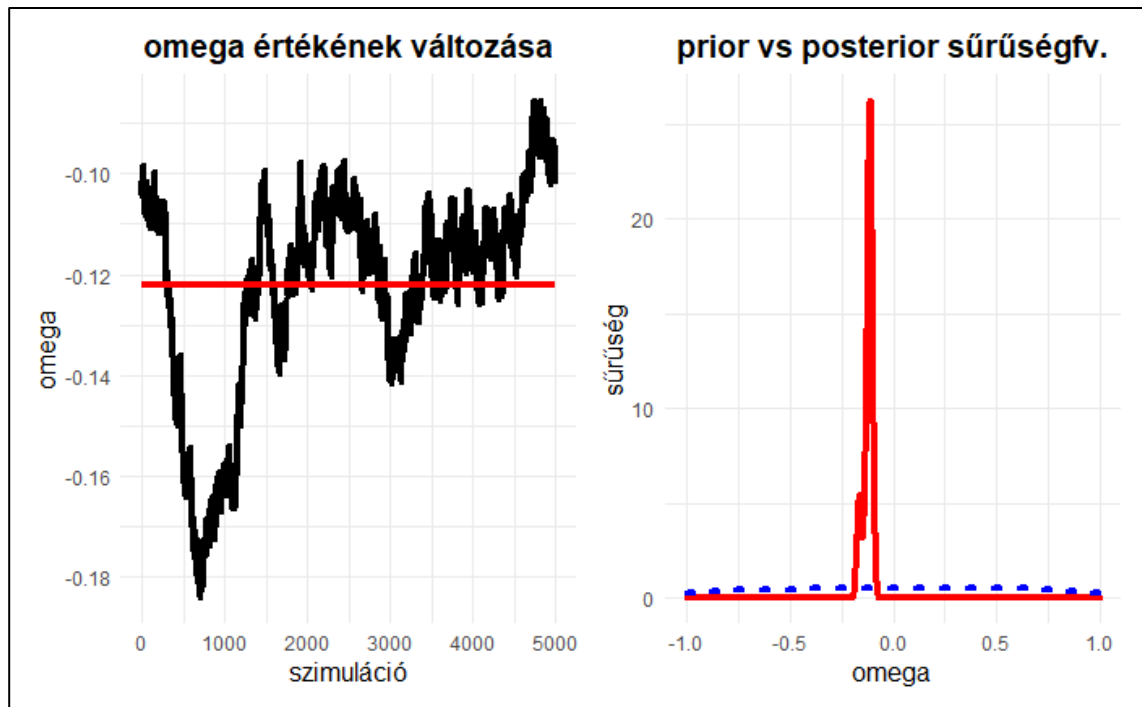
8. ábra: A ρ_2 paraméter konvergenciája és eloszlása (az észti adatsor alapján)



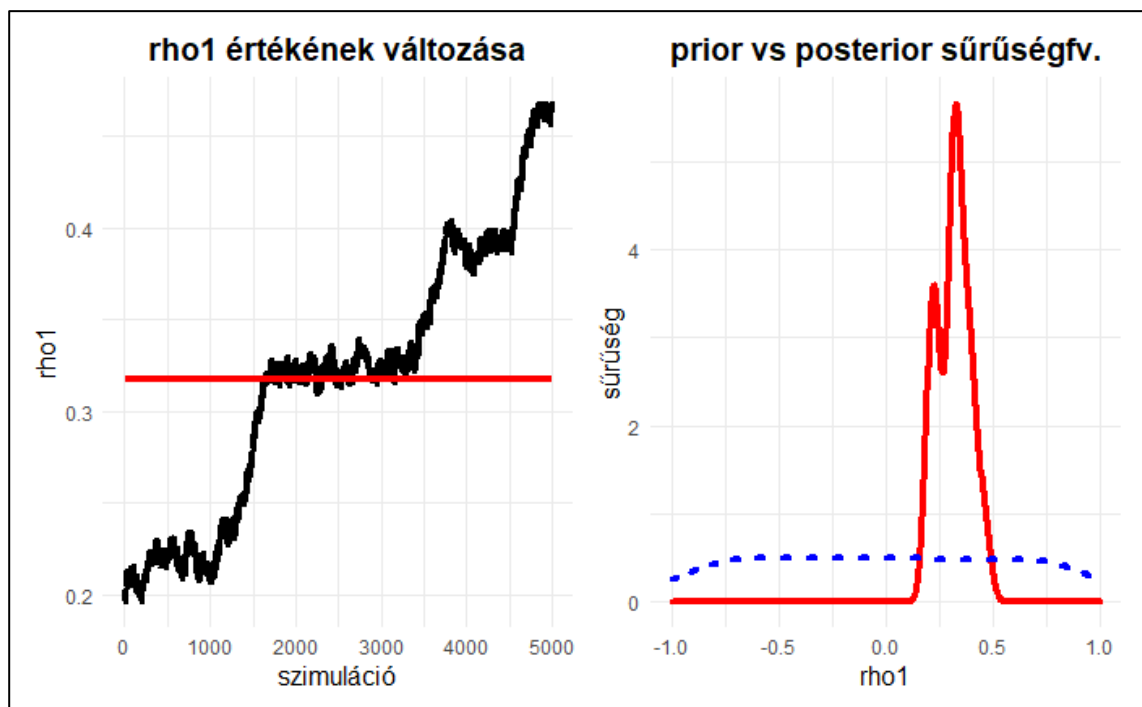
A fenti három ábrán látható az észti adatokon lefutott *MCMC* szimuláció eredménye. Sajnálatos módon az eljárás összes lehetséges hátránya megmutatkozott (többszöri futtatás után):

- i. Nullához közeli kiindulási paraméterek esetén a ρ_1 és ρ_2 paraméterek nagyon gyorsan nullához konvergálnak.
- ii. 0,4-0,5 fölötti kezdeti paraméterek esetén az eljárás hibát jelez ki, mert a *likelihood* függvény értéke olyan nagy számot vesz fel, amit az R már végtelenként kezel.
- iii. Ugyanez a hiba lép fel, ha az eljárás során az érzékenység paraméter értékét növelem meg „túlságosan” (ez a probléma már 0,5%-os potenciális lépésköz/érzékenység esetén is fellép).
- iv. A szimuláció többszöri futtatásának hatására megfigyelhetővé vált, hogy szerencsés véletlenszámgenerálás mellett úgy tűnik, mintha a folyamat konvergálna. Ez látható például ρ_2 -nél az első és nagyjából ezerötszázadik iteráció között.
- v. Az utolsó probléma, amiről szót szeretnék ejteni az a ρ_1 és a ρ_2 látszólagos konvergenciája (ezzel szemben ω , a leglényegesebb paraméter egyáltalán nem konvergál, általában a kiindulási érték körül mozog). Számos futtatás után úgy tűnik, hogy ha nagyjából 0,2 körüli kezdeti értékről 0,02-es érzékenység mellett határozzuk meg ρ_1 és ρ_2 paramétereket, akkor az előbbi egyhez tart, míg az utóbbi elég gyorsan közelít a nullához.

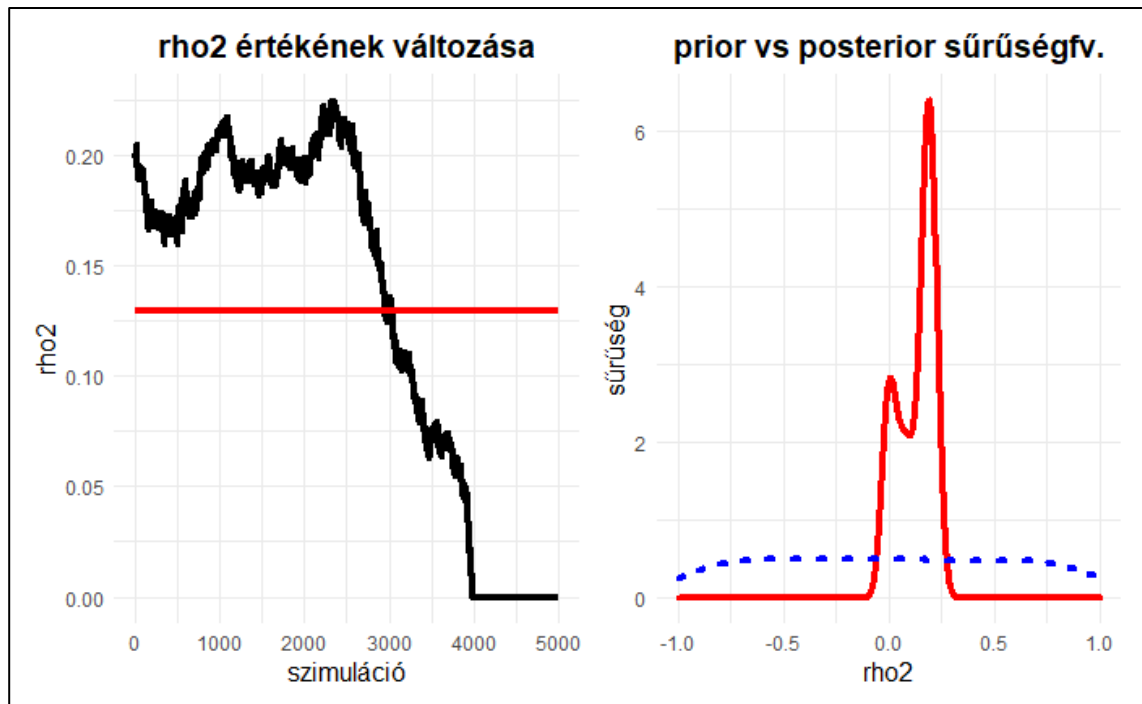
9. ábra: Az ω paraméter konvergenciája és eloszlása (a finn adatsor alapján)



10. ábra: A ρ_1 paraméter konvergenciája és eloszlása (a finn adatsor alapján)

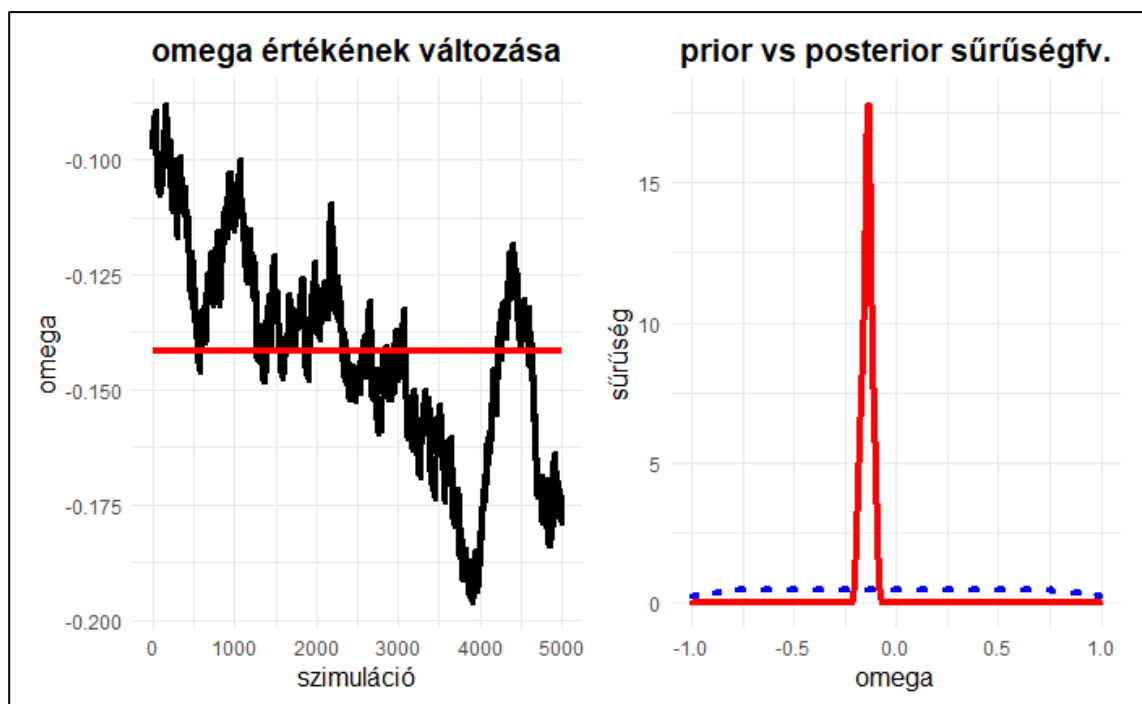


11. ábra: A ρ_2 paraméter konvergenciája és eloszlása (a finn adatsor alapján)

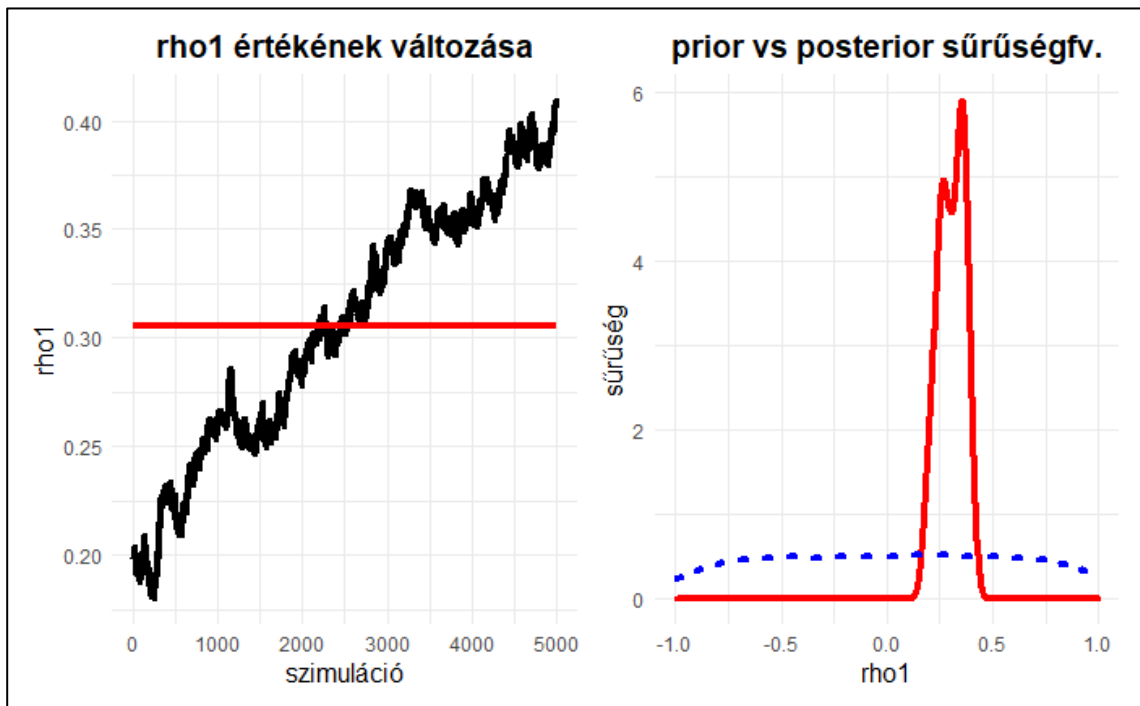


Az észrt adatoknál megfigyelt problémák láthatóak a finn adatokon végzett szimulációnál is. Ebben az esetben még egy újabb probléma is jelentkezik, mégpedig, hogy időlegesen konvergenciának tűnő szakaszok alakulnak ki (mint például a ρ_1 paraméter esetén az 1700. és az 2300. iterációs lépések között).

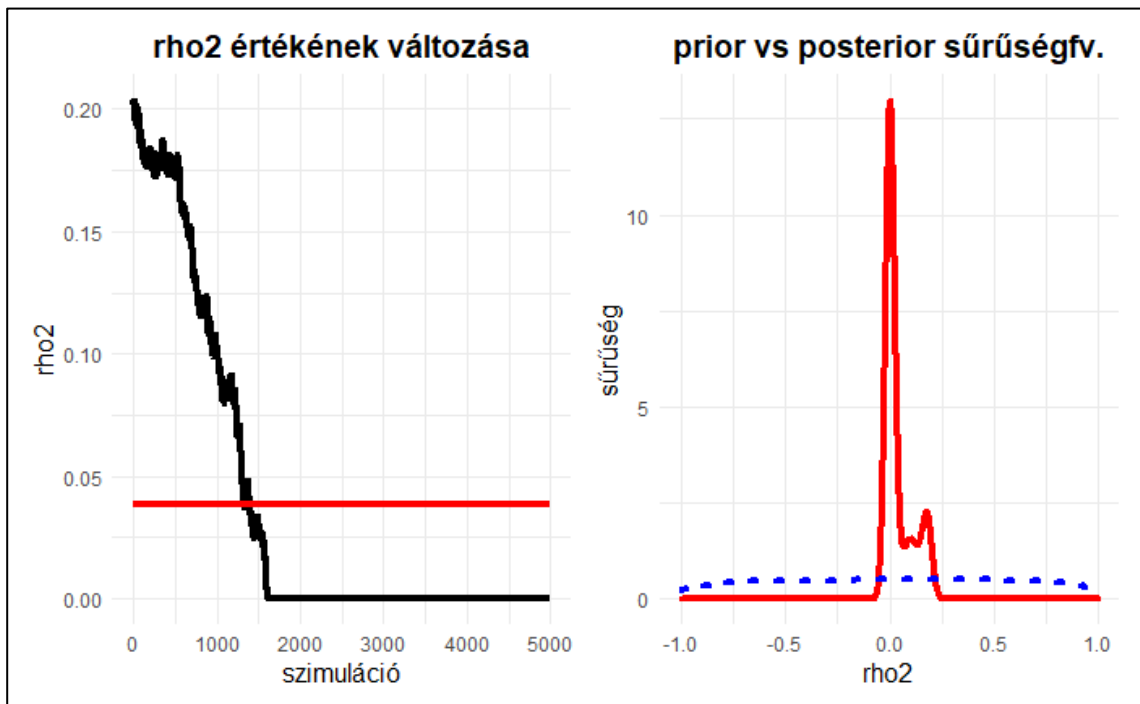
12. ábra: Az ω paraméter konvergenciája és eloszlása (a spanyol adatsor alapján)



13. ábra: A ρ_1 paraméter konvergenciája és eloszlása (a spanyol adatsor alapján)



14. ábra: A ρ_2 paraméter konvergenciája és eloszlása (a spanyol adatsor alapján)



A spanyol adatokon elvégzett szimulációnál látszik legjobban a ρ_1 és ρ_2 konvergenciája - a ρ_1 paraméter lendületesen közelít egyhez, míg a ρ_2 paraméter nagyjából 1500 iteráció után elér nagyjából nullához. Mindeközben az ω paraméterről konvergencia tekintetében semmit sem lehet elmondani.

Az általam bemutatott három és kutatásom során lefutott több száz szimuláció alapján úgy tűnik, hogy a Witzany által definiált modell nem mindig helytálló. Ezért egy modellvalidációt futtatok a cseh kutató által használt kiindulási és érzékenységi értékkel, ahol a két mögöttes faktort az elméletnek megfelelően standard normális valószínűségi változókként generálok. Megvizsgálom, hogy ekkor nekem is hasonló eredmények jönnek-e a szimulációból. A Witzany által meghatározott értékek a következők (PD és LGD korrelációs modellje):

- $\omega = 0,06$
- $\rho_1 = 0,04$
- $\rho_2 = 0,03$
- $\varepsilon_1 = 12\%$ az ω paraméterhez tartozó érzékenység
- $\varepsilon_2 = 1\%$ a ρ_1 paraméterhez tartozó érzékenység
- $\varepsilon_3 = 2\%$ a ρ_2 paraméterhez tartozó érzékenység
- $PD_0 = \frac{91\,202}{25\,572\,087} = 0,003566467$
- A két faktorra illesztett *ARMA* modellek specifikációja: $ARMA(1, 0)+ARMA(1,0)$

Kevésbé informatív eredményt kaptam. A ω paraméter néhány lépés alatt egyhez konvergált, míg a ρ_1 és a ρ_2 paraméterek is pár iteráció után nullába konvergáltak. Lecsökkentettem az érzékenységi paramétereket az általam használt (0,02-es) szintre. Ebben az esetben ω már nem tart egyhez, azonban ρ_1 és ρ_2 pillanatok alatt nullába konvergáltak. Ha pedig a paraméterek kezdőértékét 0,15 fölé viszem, akkor ugyanolyan típusú diagramokat kapok, mint amit a három ország esetén bemutattam.

Ezek alapján kijelenthető, hogy a Witzany által alkotott korrelációs modell nem minden esetben helytálló és sajnos nem lehet vele kiszámítani a PD és az RR, illetve PD és LGD között fellépő korrelációs hatásokat. A modellben egyértelműen ki kellene kristályosítani a mögöttes faktorokkal – $x_1(t)$ és $x_2(t)$ – kapcsolatos $ARMA(p, q)$ modellezés lépéseit, mert jelenleg az nagyon zavarosan van megfogalmazva (nem igazán átlátható), továbbá sokkal pontosabban kell meghatározni, hogy miképp lehet a vizsgált időszori hatást beépíteni a *likelihood* becslőfüggvénybe.

Érdeemes lehet hosszabb idősorok alapján tekinteni a modellezést, annak érdekében, hogy jobban érvényesülhessenek a mögöttes trendek és gazdasági változások. Úgy vélem, ahhoz, hogy igazán pontos modellt tudjunk alkotni nagyobb hangsúlyt kell fektetni a vigaszráta értékeit és a nemteljesítés valószínűségét alakító mögöttes folyamatok meghatározására.

Szót kell ejtenem a Witzany által alkotott modell szimulációk során tapasztalt limitációiról és a hozzá kapcsolódó módszertani aggályokról, melyet az alábbi pár pontban foglaltam össze:

- i. Az első aggály, ami megfogalmazódik bennem, a kezdeti modellel kapcsolatos. A Witzany által meghatározott kétváltozós modell matematikai értelemben egyértelmű, azonban gyakorlati implementációjával kapcsolatban problematikus kérdések merülnek föl.
- ii. Nagyon bonyolult a modell paramétereinek becslésére felírt *likelihood* függvény (mivel rengeteg változótól függ). Így a *likelihood* függvény alakítása során könnyű figyelmetlenségi hibát vétetni. A *likelihood* függvény a legkritikusabb pontja a paraméter becslésnek. Előfordulhat, hogy Witzany ebben a lépésben hibázott a modellezése során.
- iii. Ha az előző két pont során minden (matematikai értelemben) helyes és egyértelmű, akkor sem garantált, hogy a becsült paraméterek konvergálni fognak valós adatokon.
- iv. Miként már korábban is jeleztem a modellben a mögöttes faktorok időszori jellegének – az $ARMA(p, q)$ modellezésnek – az implementációja nem követhető. Számomra nem volt egyértelmű, hogy a vizsgált $ARMA$ modellek együtthatói hogyan épülnek be a modellbe és a *likelihood* függvénybe.
- v. Zárásképp pedig a mögöttes faktorok meghatározásának kérdését említem. Véleményem szerint pozitív hatással volna a modellre, ha a mögöttes faktorokat (x_1 és x_2) pénzügyi, közgazdaságtani elvek és megfontolások alapján, sokkal pontosabban határoznánk meg.

De, hogy ne csak kritikákat fogalmazzak meg Witzany modelljével kapcsolatban, fontos megemlítenem, hogy a modell nagyon jól mutatja, hogy a vigaszráta, illetve a nemteljesítés miatti veszteség modellezése mennyire bonyolult. Továbbá a modell segítségével betekintést nyerhetünk a PD és az RR korrelációs struktúrájának az empirikus működésébe.

Ez a modell még nem működik jól, azonban úgy gondolom, hogy erre lehet építeni és ezt a modellt továbbfejlesztve és pontosítva el lehet jutni egy gyakorlatban is használható *wrong-way risk* modell kifejlesztéséhez.

4.3. Regressziós modellezés (bimodalitás vizsgálata)

A módszertani fejezetben már részletesen bemutattam a vigaszráta és a nemteljesítés miatti veszteség eloszlását becslő regressziós módszereket és az ehhez kapcsolódó szakirodalmakat. Most ezen eljárások közül hármat – lineáris regressziót, *GLM* (általánosított lineáris) regressziót és béta regressziót – fogok alkalmazni a saját adatbázisomon. Minden becslést négy adatkészleten – a teljes adatbázison, illetve az észt, a finn és a spanyol adatsorokon – végzek el, így különböző jellegű adatsokaságok alapján tudom tesztelni az eljárásokat. Ehhez a 3. fejezetben bemutatott változókat fogom használni a Bondora adatbázisból (transzformálva). A változószelekcióhoz (kezdeti regresszióhoz) az alábbi változókat használtam föl:

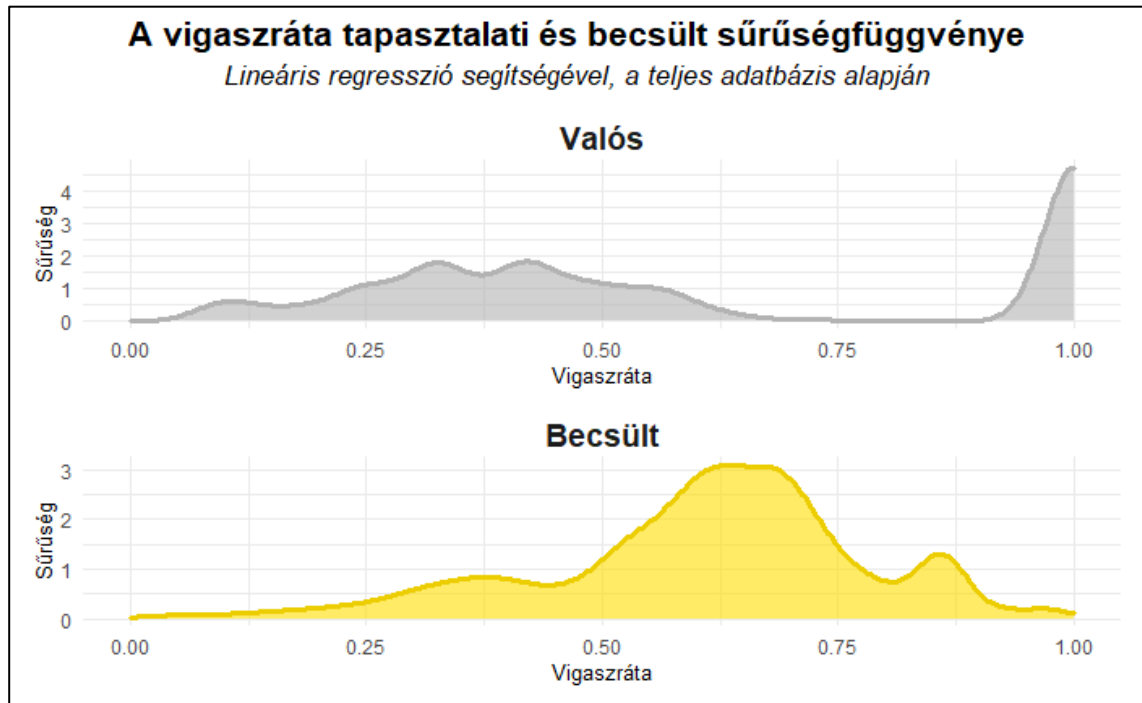
- *PD*
- *EAD*
- *Age*
- *D_woman*
- *D_undef*
- *D_irodai*
- *D_egyeb_munka*
- *Education*
- *IncomeTotal*
- *MontlyPayment*
- *Debt_to_Income*

E változók közül 5%-on az *Education* és az *IncomeTotal* nem volt szignifikáns, ezért ezt a kettőt kihagytam a további modellezésből. Ugyanerre a konklúzióra jutottam forward és backward változószelekcióval is. Az *IncomeTotal*-t kihagyva az *Education* éppen szignifikánssá válna, de azért hagytam ki, mert a modell magyarázóerejét (R^2) mindösszesen 0,0001-gyel növelte, de cserében plusz egy változót jelente, ami problematikussá válhat a nagy számításigényű bonyolult eljárások, mint például a béta regresszió lefuttatásakor.

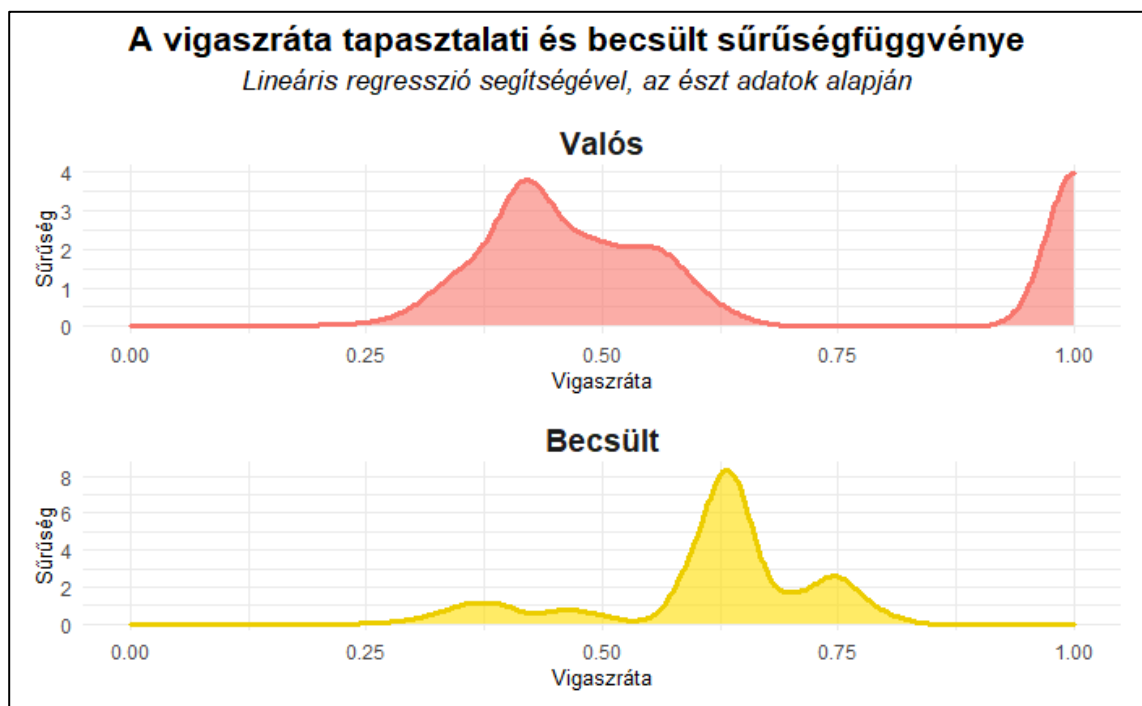
Először a lineáris regresszió (*OLS*) segítségével történő becslést mutatom be, ezt követően áttérek az általánosított lineáris regresszióra, végül pedig a béta regresszióra. Azért ezt a három módszert választottam, mert ezek a legegyszerűbbek és ezek segítségével be tudom mutatni, hogy hogyan működik a valós sűrűségfüggvény (ezáltal az eloszlás) becslése regressziós eljárások segítségével. Az ennél bonyolultabb, összetett módszerek is hasonlóan működnek,

azonban azok sokkal több számítási kapacitást emésztenek fel. A célokat pedig egyszerűbb regressziós módszerekkel is el tudom érni.

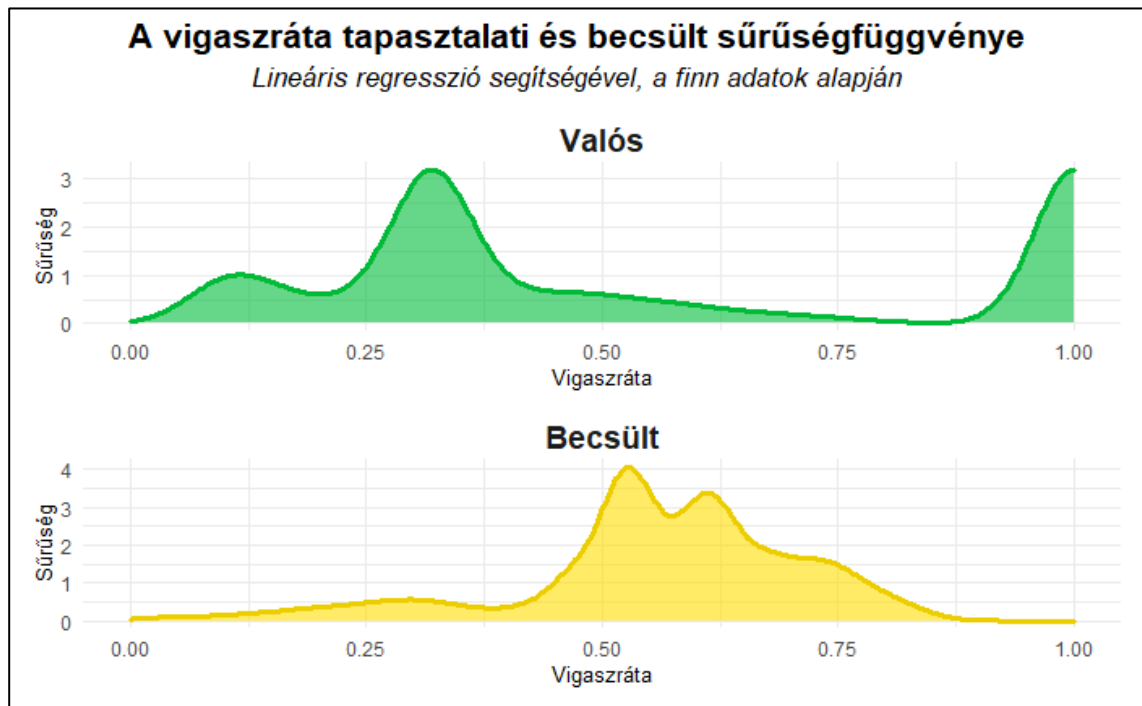
15. ábra: Vigaszráta eloszlásának becslése lineáris regresszióval (teljes adatbázis alapján)



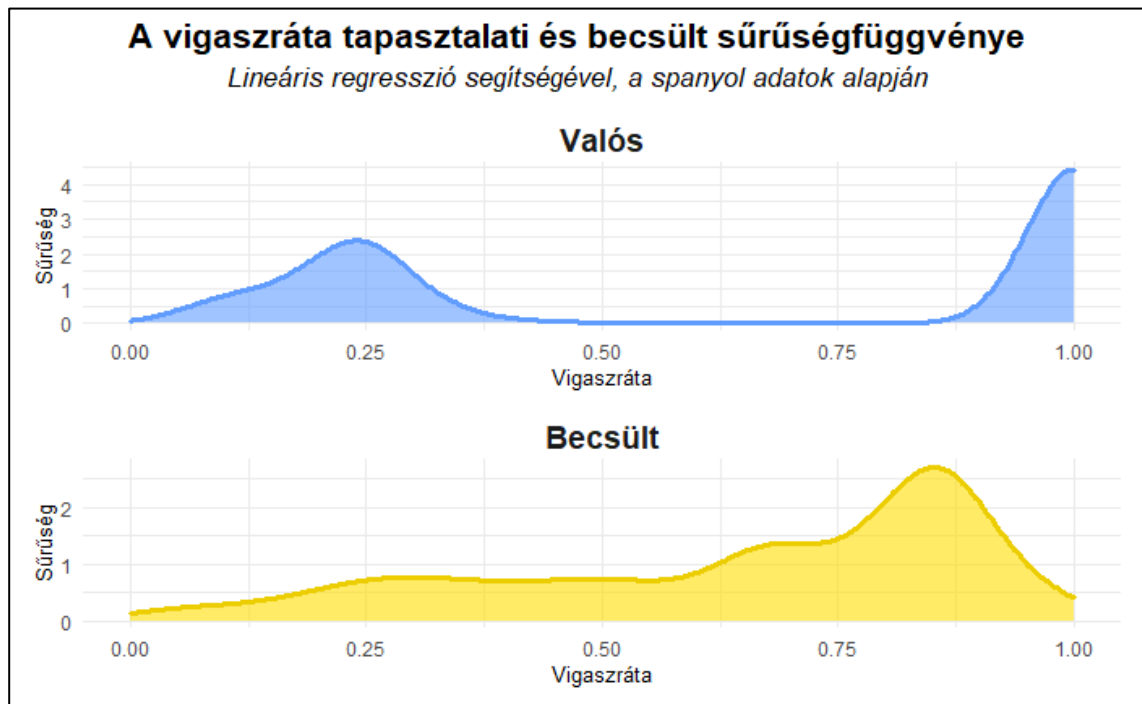
16. ábra: Vigaszráta eloszlásának becslése lineáris regresszióval (észti adatok alapján)



17. ábra: Vigaszráta eloszlásának becslése lineáris regresszióval (finn adatok alapján)



18. ábra: Vigaszráta eloszlásának becslése lineáris regresszióval (spanyol adatok alapján)

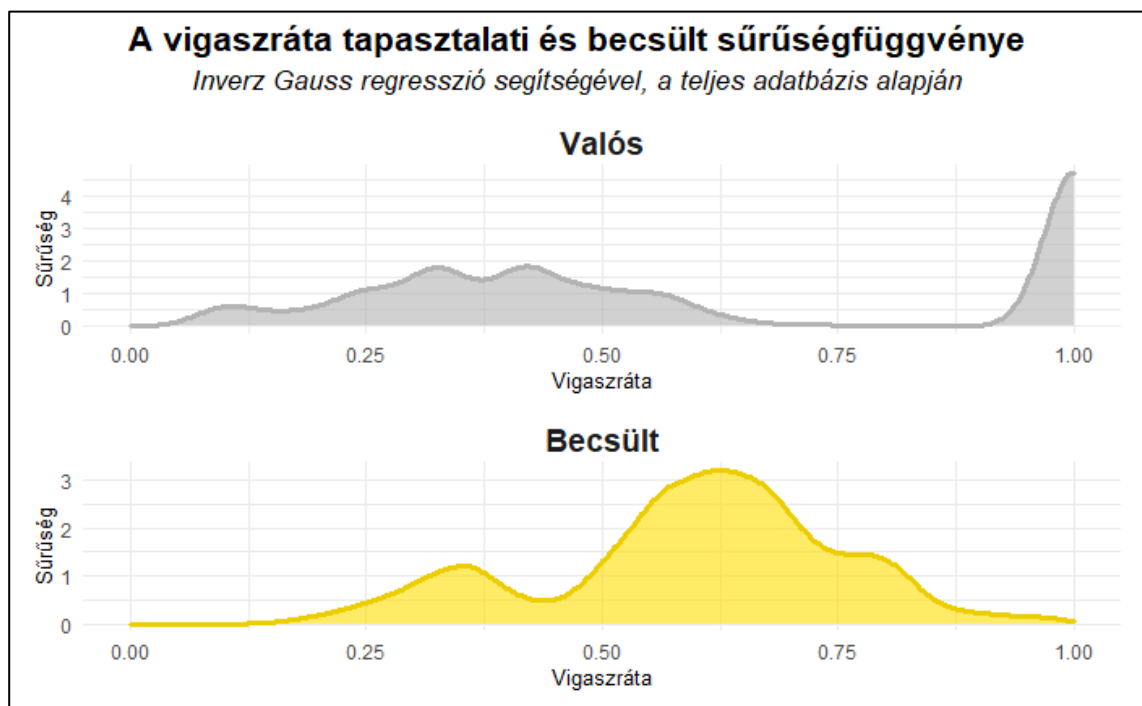


A lineáris regresszióval közepesen jó mintán belüli becslést értem el. Az egyes modellekhez tartozó R^2 -ek értéke 0,20 és 0,56 között mozgott, az $RMSE$ pedig 0,23 és 0,29 között vett föl értékeket. Azonban az eloszlás becslése – ahogy azt a módszertani fejezetben bemutattam – elég gyenge lett, kevésbé illeszkedik a valós eloszlásra (sűrűségfüggvényre).

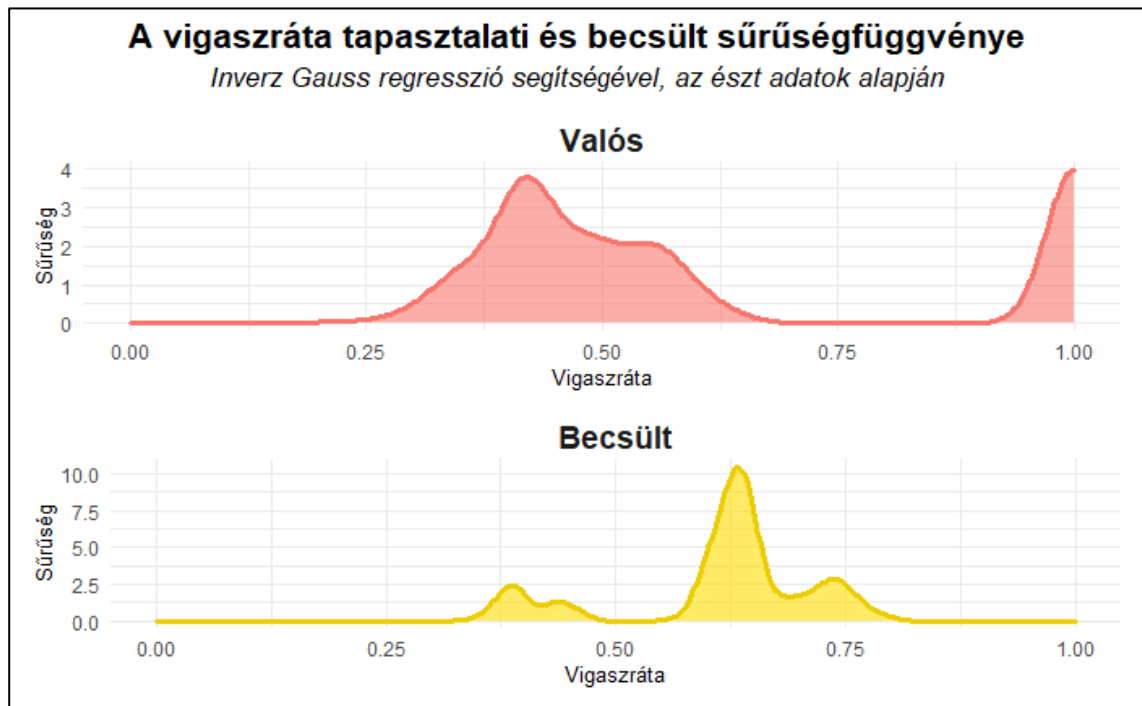
Egyik esetben sem rendelkezik bimodális tulajdonsággal az illesztett kernel típusú sűrűségfüggvény. Továbbá négy esetből háromnál az $x = 1$ határpont környékén található sűrűsödést nem tudta előrejelezni a becslés, ami jelentős probléma.

A második regressziós módszer, amit használok az általánosított lineáris regressziós (transzformációs) módszerek csoportjába tartozó inverz Gauss regresszió (IGR), erről már szót ejtettem a módszertani fejezetben. Ebben az esetben a linkfüggvény – amely a folyamatot lineárisra transzformálja – a logaritmusfüggvény (\log) volt. Nagy előnye a módszernek, hogy az OLS regresszióval becsült értékeket az inverzfüggvény visszatranszformálja a $[0, 1]$ zárt intervallumra, így minden becslés a számunkra megfelelő intervallumba fog esni.

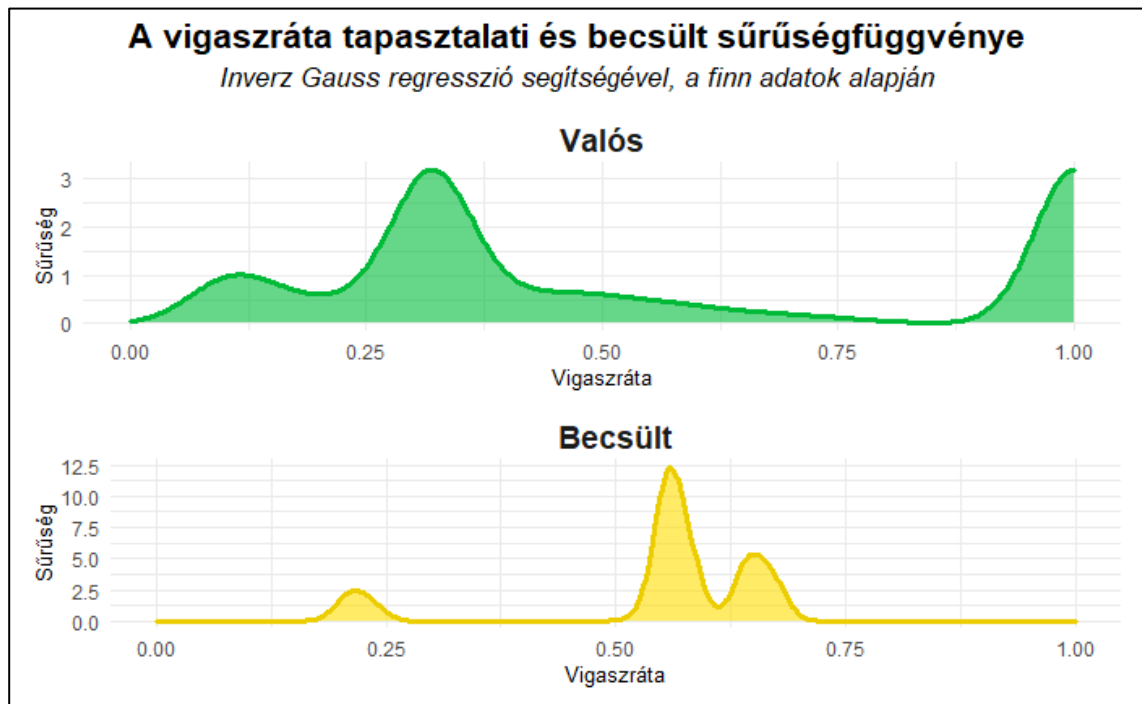
19. ábra: Vigaszráta eloszlásának becslése GLM regresszióval (teljes adatbázis alapján)



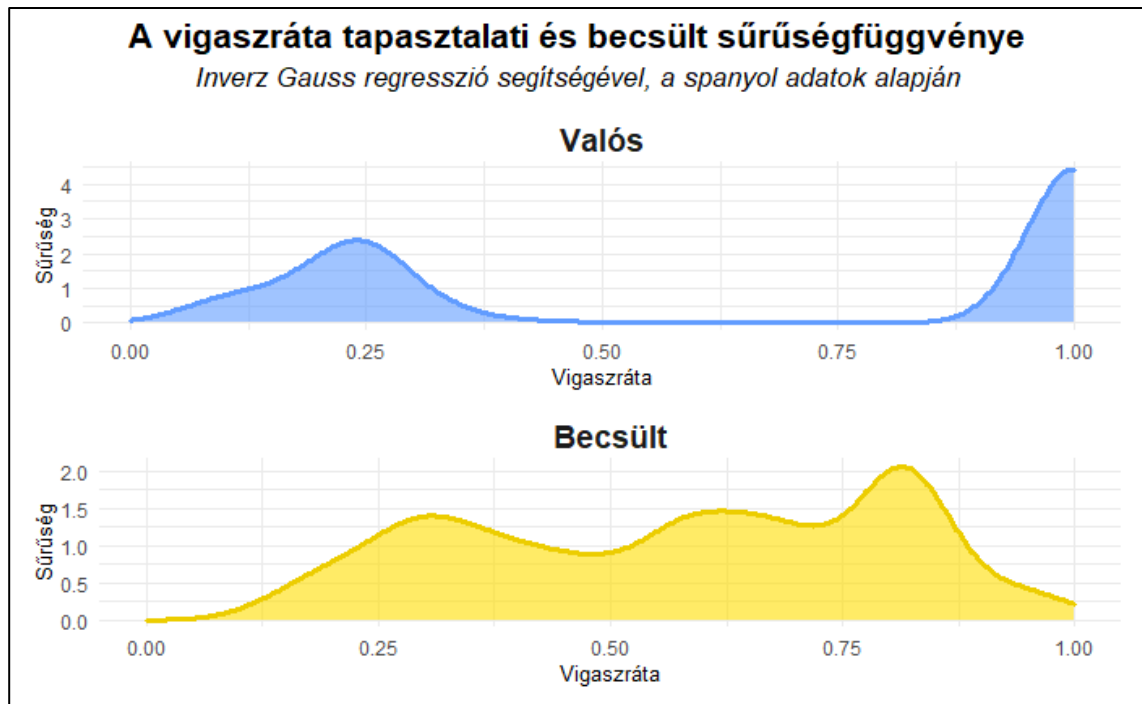
20. ábra: Vigaszrása eloszlásának becslése GLM regresszióval (észr adatok alapján)



21. ábra: Vigaszrása eloszlásának becslése GLM regresszióval (finn adatok alapján)



22. ábra: Vigaszráta eloszlásának becslése GLM regresszióval (spanyol adatok alapján)



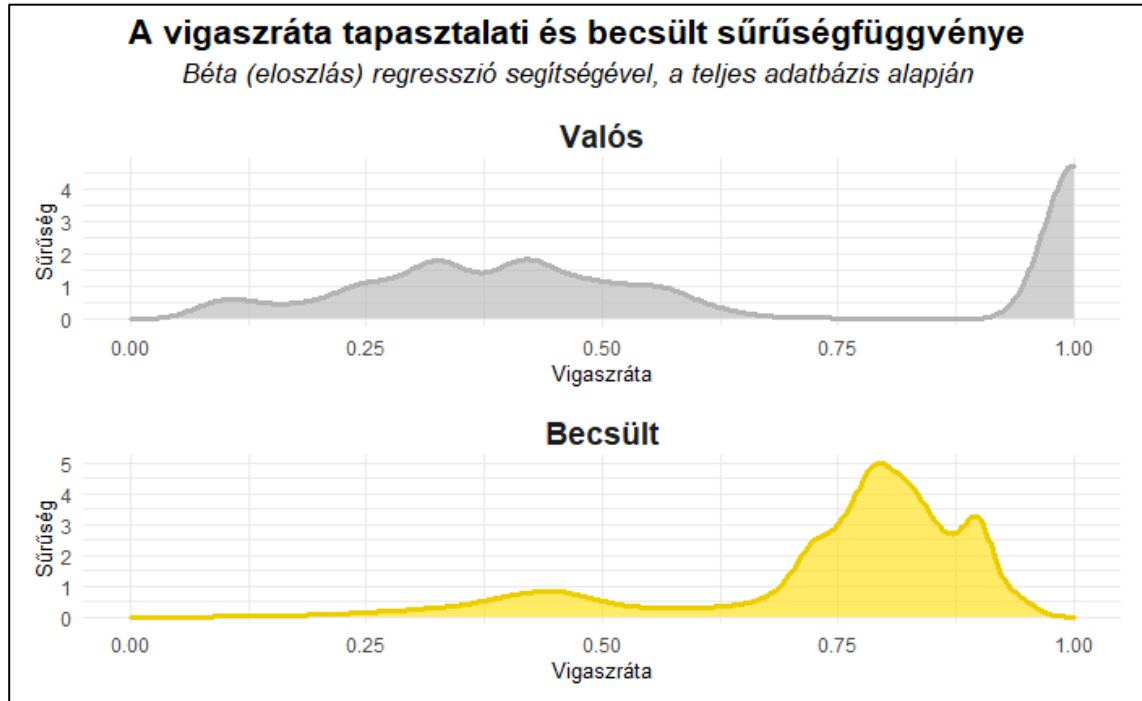
Az inverz Gauss regresszióval is közepesen jó mintán belüli becslést értem el. Az egyes modellekhez tartozó R^2 -ek értéke 0,21 és 0,56 között mozgott, az RMSE pedig 0,44 és 0,94 között vett föl értékeket. Azonban az eloszlás becslése rosszabb lett, mint az OLS regresszió esetén, bár ez összhangban van a bemutatott tanulmányok eredményeivel.

A négy eset közül egyiknél sem sikerült az egyhez közeli vigaszrátákat pontosan megbecsülni, ezáltal az $x = 1$ határpont környékén található sűrűsödést nem tudta előrejelezni az általánosított lineáris regresszió alapuló inverz Gauss módszer sem. További probléma, hogy a vigaszráta (valóságos) bimodális jellege sem jelenik meg. Bár mind a négy eset produkál valamilyen bimodális vagy trimodális jelleget, csak nem feltétlenül azokon a pontokon, ahol kellene. Ezáltal a valós és becült kernel típusú sűrűségfüggvények nem fedik egymást. Sajnálatosan ezen regressziós eljárás esetében sem sikerült a valóságoshoz közeli sűrűségfüggvényt becüslni – még annyira sem, mint amennyire a bemutatott publikációk esetén sikerült – azonban a módszertan és az általánosított lineáris regresszió bemutatása szempontjából sikeresnek ítélem meg a vizsgálódásaimat.

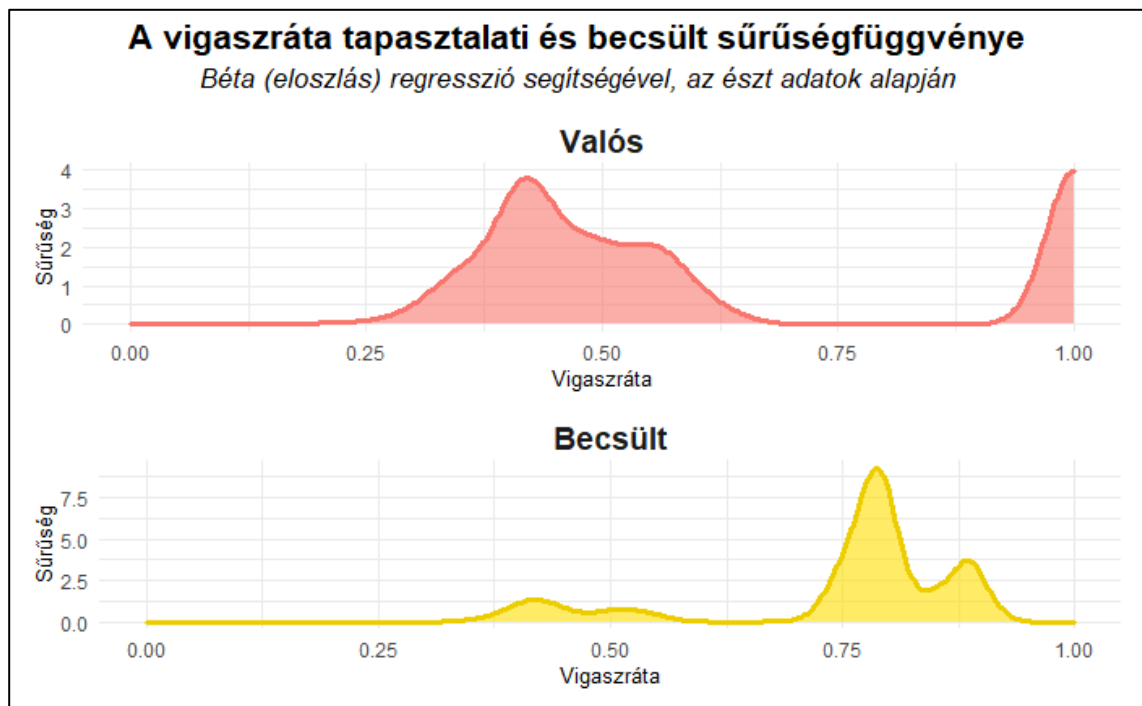
A harmadik és egyben utolsó regressziós módszer, amelyet gyakorlatban is alkalmaztam a béta regresszió. Ennek az eljárásnak az a különlegessége, hogy csak (0, 1) közötti nyílt intervallumon van értelmezve, ezért a két határpont epsilon sugarú környezetében található vigaszráta értékeket $(0 + \varepsilon)$ és $(1 - \varepsilon)$ értékekké transzformáltam, ε -t pedig 0,000001-nek

választottam. Ebben az esetben *complementary log-log (cloglog)* linkfüggvényt használtam a regressziós probléma linearizálására (ez jelentősen meggyorsította az algoritmus futását).

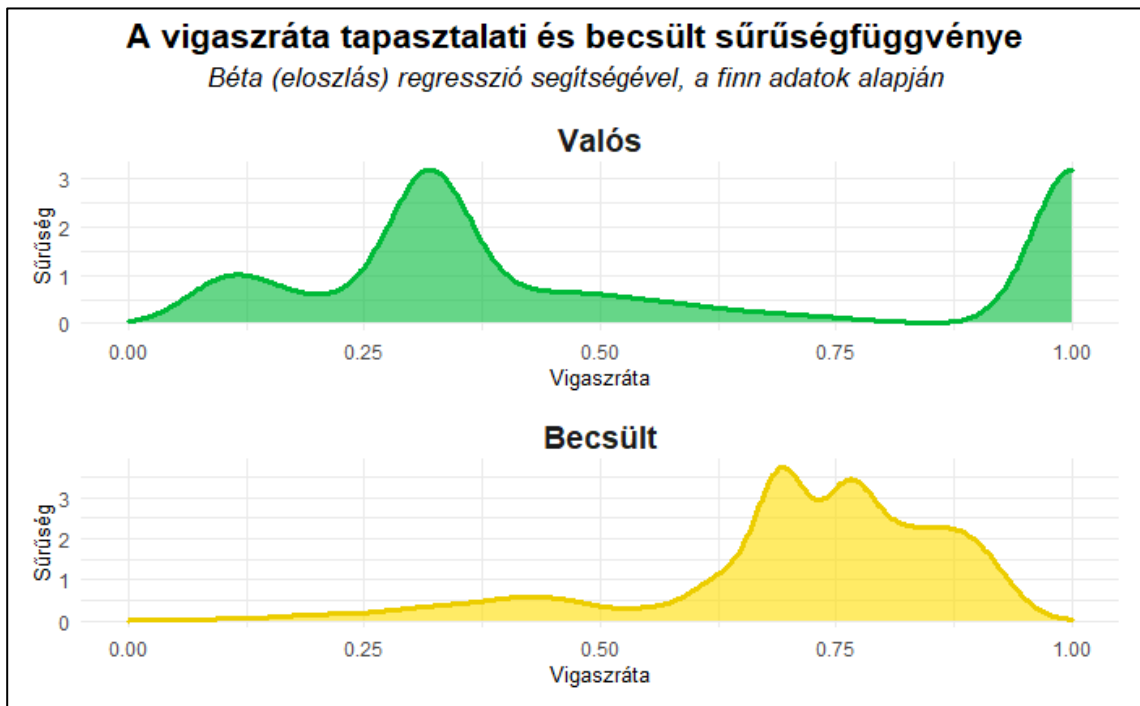
23. ábra: Vigaszráta eloszlásának becslése béta regresszióval (teljes adatbázis alapján)



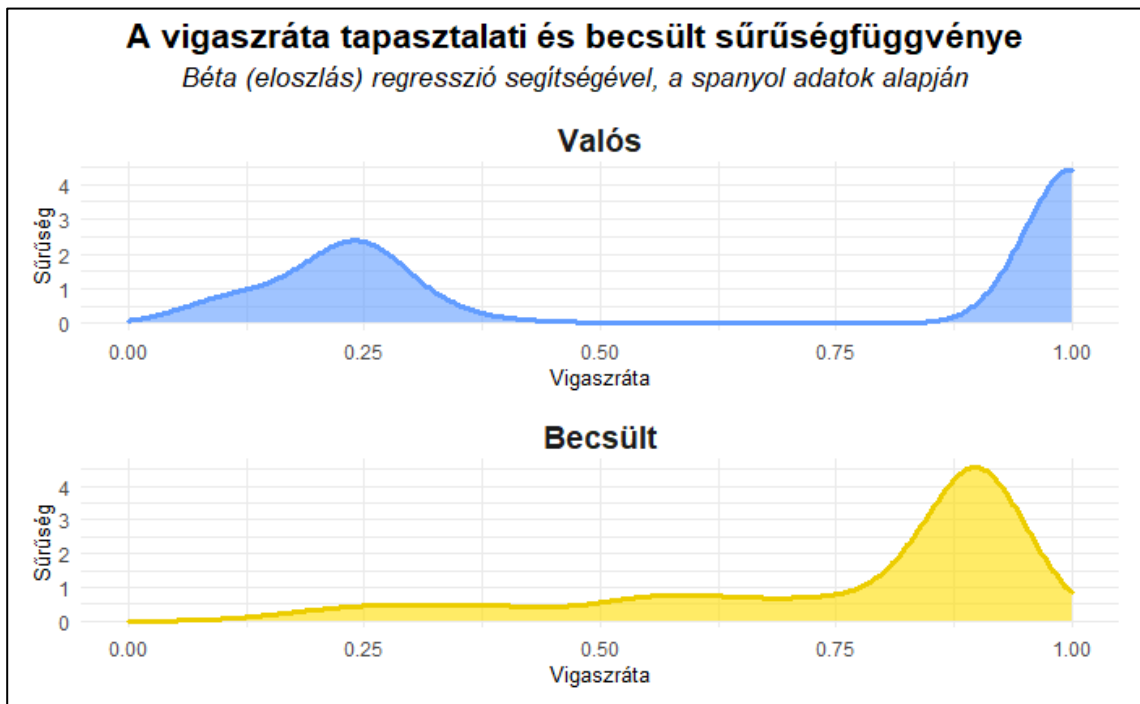
24. ábra: Vigaszráta eloszlásának becslése béta regresszióval (észti adatok alapján)



25. ábra: Vigaszráta eloszlásának becslése béta regresszióval (finn adatok alapján)



26. ábra: Vigaszráta eloszlásának becslése béta regresszióval (spanyol adatok alapján)



A bétaregresszióval is közepesen jó mintán belüli becslést értem el. Az egyes modellekhez tartozó R^2 -ek értéke 0,19 és 0,53 között mozgott, az RMSE pedig 0,27 és 0,33 között vett föl értékeket. Az eljárás a teljesítménymérő számok alapján kicsit gyengébb, mint az OLS regresszió, azonban két nagy előnye van vele szemben. Egyrészt a becsült sűrűségfüggvény már jobban fedi a valós sűrűségfüggvényt, másrészt $x = 1$ határponthoz közeli értékeket pontosabban közelíti. Továbbá a másik két regressziós módszerrel szemben, itt már megjelenik a bimodális jelleg, igaz nem olyan szépen, mint ahogy szeretném, de így is jobb eredményt mutatva, mint amennyit a másik két eljárással elértem.

Bonyolultabb modellek segítségével várhatóan jobb eredményeket lehet elérni és pontosabb becsléseket adni mind a vigaszráta értékei, mind pedig az RR eloszlására. De még a bemutatott modelleket is számos módon lehet javítani:

- i. nagyobb adatbázist alapul véve (kevésbé diszkrét példa),
- ii. több magyarázó változót vinni be a modellekbe (pontosabb és részletesebb magyarázat),
- iii. változók közti nemlineáris kapcsolatokkal is számolni (pl. interakciók),
- iv. mintán kívüli előrejelzés alkalmazása (pl. keresztvalidáció)

Azonban a pontosabb modelleknek is van „ára”, mégpedig, hogy a modellek egyre inkább bonyolultabbá válnak és ezáltal egyre számításigényesebbek lesznek a becslések.

5. Összefoglalás, konklúzió

A szakdolgozatom célja egy módszertani áttekintő nyújtása a vigaszráta és a nemteljesítés miatti veszteség modellezéséről. Ehhez kapcsolódóan bemutattam a fő (üzleti) modellezési kihívásokat:

- I. A rosszirány kockázatot (*wrong-way risk*), ami az RR (illetve LGD) és a PD közt fennálló (lineáris korrelációs) kapcsolatból származik.
- II. A vigaszráták bimodális (vagy multimodalitás) tulajdonságának reprodukálásának a nehézségeit.

A 2. fejezetben áttekintettem, hogy milyen módszertani fejlődés előzte meg a legújabb vigaszrátával és nemteljesítés miatti veszteséggel foglalkozó modellek kialakulását. Ezt követően részletesen bemutattam a két modellezési területhez kapcsolódó gyakorlatorientált módszereket az utóbbi húsz év főbb publikációinak ismertetésével. A módszertanokat ismertető alfejezetek végén készítettem egy-egy összefoglaló táblázatot arról, amely áttekinti a vizsgált modelleket és eljárásokat, azok főbb eredményeit, erősségeit és esetlegesen hátrányait. A fejezet végén sor kerül a két módszertan összehasonlítására is. Úgy vélem, jelenleg még nem összekapcsolható a két modellezési módszertan, azonban pár éven belül, a terület fejlődésének hatására a két módszertant kombinálni lehet.

A korrelációs (*wrong-way risk*) modellek matematikai szerkezetét erősen befolyásolja, hogy a kutatók számára milyen típusú adatok állnak rendelkezésre. Ezért két különböző szerkezetű (alapvető szemléletbeli különbséggel rendelkező) gyakorlatorientált *wrong-way risk* modellt mutattam be. Witzany modelljéhez aggregált adatokra van szükség, Bade, Rösch és Scheule modelljéhez pedig létfontosságúak az olyan számlaszintű egyedi adatok, amelyekhez napi rendszerességű kereskedési információk (például értékváltozás) révén juthatunk el.

A módszertani fejezetem második felében áttekintettem hét olyan publikációt, amelyek a vigaszráta vagy a nemteljesítés miatti veszteség becslésével, valamint az RR és az LGD eloszlásának bimodalitásával (egyres esetekben multimodalitásával) foglalkoznak. A tanulmányok segítségével bemutattam, hogyan próbálnak a kutatók egyre precízebb és valósághűbb modelleket alkotni, annak érdekében, hogy minél pontosabb lehessen megbecsülni a vigaszrátát és annak eloszlást. Az ismertetett modellek közül kiválasztottam hármát, amelyek gyakorlati alkalmazását is bemutattam.

Kutatásomhoz a Bondora adatbázisát használtam föl, amelyben számlaszintű adatok találhatóak. Az adatállomány „stock” típusú adatokból áll és a portfólió nem kereskedett napi szinten (azaz nem áll rendelkezésre napi áringadozásról információ), ezért csak Witzany modelljét tudtam alkalmazni. Ezzel szemben regressziós modellezéshez ideális a Bondora adatbázisa, mert az egyedi hitelekhez kapcsolódóan rengeteg (szocioökonometriai) információ elérhető, amelyeket fel lehet használni magyarázó változóként a modellezés során.

Az empirikus kutatásom során beláttam Witzany modelljéről, hogy nem mindig működik. Ennek ellenére nagyon jó volt arra, hogy bemutassam, hogyan lehet építőkövekként felépíteni egy konkrét korrelációs modellt, mik ennek az előnyei, nehézségei és limitációi. Továbbá a modell segítségével áttekintettem a módszertanhoz kapcsolódó gyakorlati eljárásokat. Így például levezettem a paraméterbecsléshez tartozó *likelihood* függvényt és bemutattam a bayes-i *MCMC* szimulációt, amely egy nagyon hasznos eszköz numerikus optimalizációra.

A gyakorlatban is bemutatott három regressziós eljárás – lineáris regresszió (*OLS*), invez Gauss regresszió (*IGR*) és béta regresszió – közül egyik sem adott igazán jó becslést a vigaszráta értékére és nem is tudta visszaadni annak bimodális jellegét. Ennek ellenére úgy vélem a módszertan tulajdonságait, nehézségeit és limitációt be tudtam mutatni. Az elméleti és gyakorlati fejezetek áttekintése után arra a következtetésre jutottam, hogy a regressziós eljárások során több szocioökonometriai változót és komplexebb modelleket kell felhasználni, a pontosabb (eloszlás) becslés érdekében, még annak árán is, hogy az egyre bonyolultabb eljárások igen számításigényesek. De nem feledkezhetünk meg Qi és Zhao eredményéről sem, mely szerint az erős bimodális jelleg még nem jelent jó modellilleszkedést (Qi & Zhao, 2011).

A kutatásom során számos fontos tanulság fogalmazódott meg bennem, ezeket az alábbi pár pontban foglaltam össze:

- i. A vigaszráták viselkedésének kutatása egy nagyon összetett terület, számos módszertan együttes alkalmazása szükséges a vizsgálatához.
- ii. Az ehhez kapcsolódó modellezés nagyon adat és számításigényes.
- iii. Fontos tanulság volt számomra, hogy nem mindig a legkomplexebb modellek és eljárások működnek a legjobban.
- iv. Kritikusan kell állni minden modellhez, mindaddig amíg be nem látjuk, hogy azok helytállóak a valós adatokon (empirikus kutatás jelentősége) nyugvó gyakorlatban is.
- v. A vigaszrátákhoz kapcsolódó modellek tanulmányozása és fejlesztése nagyon fontos, annak érdekében, hogy a pénzügyi szolgáltatók pontosabban ki tudják számolni tőkekövetelményeiket (gyakorlati szempont).

A szakdolgozatomhoz kapcsolódó kutatásom során két fontos, de a dolgozatom keretein túlmutató kérdés is megfogalmazódott bennem:

- A. A két módszertan kombinációjával előre lehet-e jelezni nemteljesítéskor a később realizálódó vigaszrátákat?
- B. Fel lehet-e használni e két módszert a behajtási folyamatok modellezésére?

A dolgozat végére érve kijelenthetem, hogy a vigaszráta modellezéséhez kapcsoló módszerek és eljárások kérdésének vizsgálatával egy nagyon érdekes témakört sikerült feldolgoznom és reményeim szerint bemutatnom.

Hivatkozásjegyzék

- Altman, E., Resti, A., & Sironi, A. (2004). Default Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence. *Economic Notes*, 33(2), 183–208. <https://doi.org/10.1111/j.0391-5026.2004.00129.x>
- Bade, B., Rösch, D., & Scheule, H. (2011). Default and Recovery Risk Dependencies in a Simple Credit Risk Model. *European Financial Management*, 17(1), 120–144. <https://doi.org/10.1111/j.1468-036X.2010.00582.x>
- Basel Committee on Banking Supervision. (2005). *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. <https://www.bis.org/publ/bcbs118.htm>
- Bellotti, T., & Crook, J. (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting*, 28(1), 171–182. <https://doi.org/10.1016/j.ijforecast.2010.08.005>
- Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Calabrese, R. (2014). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry*, 30(2), 99–114. <https://doi.org/10.1002/asmb.1932>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chava, S., Stefanescu, C., & Turnbull, S. (2011). Modeling the Loss Distribution. *Management Science*, 57(7), 1267–1287. <https://doi.org/10.1287/mnsc.1110.1345>
- Credit Suisse Financial Products. (1997). *CreditRisk+ A Credit Risk Management Framework*. http://www.defaultrisk.com/pp_model_21.htm
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. John Wiley & Sons.
- Duffie, D. (1998). Defaultable Term Structure Models with Fractional Recovery of Par. *Undefined*. <https://www.semanticscholar.org/paper/Defaultable-Term-Structure-Models-with-Fractional-Duffie/5ecc7bcc595362c57b47a1d1f3ed564291d4c377>
- Duffie, D., & Singleton, K. J. (1999). Modeling Term Structures of Defaultable Bonds. *The Review of Financial Studies*, 12(4), 687–720.

- Frye, J., Ashley, L., Bliss, R., Cahill, R., Calem, P., Foss, M., Gordy, M., Jones, D., Lemieux, C., Lesiak, M., Lobbes, C., McGrew, L., Mehta, P., Nelson, J., & Waggoner, E. (2000). Collateral damage: A source of systematic credit risk. *Risk Magazine*, 13(4), 91–94.
- Frye, J. H. (2000). Depressing Recoveries. *Risk Magazine*, 13(11), 108–111.
- Gupton, G., Finger, C., & Bhatia, M. (1997). *CreditMetrics—Technical Document*.
- Hull, J., & White, A. (1995). The impact of default risk on the prices of options and other derivative securities. *Journal of Banking & Finance*, 19(2), 299–322.
[https://doi.org/10.1016/0378-4266\(94\)00050-D](https://doi.org/10.1016/0378-4266(94)00050-D)
- Jarrow, R. A., Lando, D., & Turnbull, S. M. (1997). A Markov Model for the Term Structure of Credit Risk Spreads. *The Review of Financial Studies*, 10(2), 481–523.
<https://doi.org/10.1093/rfs/10.2.481>
- Jarrow, R. A., & Turnbull, S. M. (1995). Pricing Derivatives on Financial Securities Subject to Credit Risk. *The Journal of Finance*, 50(1), 53–85. <https://doi.org/10.2307/2329239>
- Johannes, M., & Polson, N. (2010). CHAPTER 13—MCMC Methods for Continuous-Time Financial Econometrics. In Y. Aït-sahalia & L. P. Hansen (Szerk.), *Handbook of Financial Econometrics: Applications* (Köt. 2, o. 1–72). Elsevier.
<https://doi.org/10.1016/B978-0-444-53548-1.50003-9>
- Kellner, R., Nagl, M., & Rösch, D. (2021). Opening the black box – Quantile neural networks for loss given default prediction. *Journal of Banking & Finance*, 134, 106334.
<https://doi.org/10.1016/j.jbankfin.2021.106334>
- Kim, I. J., Ramaswamy, K., & Sundaresan, S. (1993). Does Default Risk in Coupons Affect the Valuation of Corporate Bonds?: A Contingent Claims Model. *Financial Management*, 22(3), 117–131. <https://doi.org/10.2307/3665932>
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679.
<https://doi.org/10.1016/j.ijforecast.2015.12.003>
- Krüger, S., & Rösch, D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking & Finance*, 79, 42–56. <https://doi.org/10.1016/j.jbankfin.2017.03.001>
- Lando, D. (1998). On cox processes and credit risky securities. *Review of Derivatives Research*, 2(2), 99–120. <https://doi.org/10.1007/BF01531332>
- Litterman, R. B., & Iben, T. (1991). Corporate bond valuation and the term structure of credit spreads. *The Journal of Portfolio Management*, 17(3), 52–64.
<https://doi.org/10.3905/jpm.1991.409331>

- Longstaff, F. A., & Schwartz, E. S. (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *The Journal of Finance*, 50(3), 789–819.
<https://doi.org/10.2307/2329288>
- Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, 28(1), 161–170. <https://doi.org/10.1016/j.ijforecast.2011.01.006>
- Lynch, S. M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. <https://link.springer.com/book/10.1007/978-0-387-71265-9>
- Madan, D. B., & Unal, H. (1998). Pricing the risks of default. *Review of Derivatives Research*, 2(2), 121–160. <https://doi.org/10.1007/BF01531333>
- Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2), 449–470. <https://doi.org/10.2307/2978814>
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619–632. [https://doi.org/10.1002/\(SICI\)1099-1255\(199611\)11:6<619::AID-JAE418>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1)
- Pykhtin, M. (2003). Unexpected recovery risk. *Risk Magazine*, 16(1), 74–78.
- Qi, M., & Zhao, X. (2011). Comparison of modeling methods for Loss Given Default. *Journal of Banking & Finance*, 35(11), 2842–2855.
<https://doi.org/10.1016/j.jbankfin.2011.03.011>
- Siao, J.-S., Hwang, R.-C., & Chu, C.-K. (2015). Predicting recovery rates using logistic quantile regression with bounded outcomes. *Quantitative Finance*, 16(5), 777–792.
<https://doi.org/10.1080/14697688.2015.1059952>
- Tasche, D. (2004). *The Single Risk Factor Approach to Capital Charges in Case of Correlated Loss Given Default Rates* (SSRN Scholarly Paper ID 510982). Social Science Research Network. <https://doi.org/10.2139/ssrn.510982>
- Wilson, T. C. (1998). Portfolio credit risk. *Economic Policy Review*, 4(Oct), 71–82.
- Witzany, J. (2008). *Basel II Capital Requirement Sensitivity to the Definition of Default* (SSRN Scholarly Paper ID 1274186). Social Science Research Network.
<https://doi.org/10.2139/ssrn.1274186>
- Witzany, J. (2009). *Estimating LGD Correlation* (SSRN Scholarly Paper ID 1410070). Social Science Research Network. <https://doi.org/10.2139/ssrn.1410070>
- Witzany, J. (2011). *A Two-Factor Model for PD and LGD Correlation*.
<https://doi.org/10.2139/ssrn.1476305>

Ye, H., & Bellotti, A. (2019). Modelling Recovery Rates for Non-Performing Loans. *Risks*, 7(1), 19. <https://doi.org/10.3390/risks7010019>

Adatforrások jegyzéke

Bondora: Public Reports. *Loan Dataset* (Letöltés ideje: 2022. március 17., helye: <https://www.bondora.com/hu/public-reports#shared-legend>)

Eurostat: *Main GDP aggregates per capita* (Letöltés ideje: 2022. március 17., helye: https://ec.europa.eu/eurostat/web/products-datasets/-/namq_10_pc)

Eurostat: *Unemployment rate - quarterly data, seasonally adjusted* (Letöltés ideje: 2022. március 17., helye: <https://ec.europa.eu/eurostat/web/products-datasets/-/tipsun30>)

Eurostat: *HICP - monthly data (index)* (Letöltés ideje: 2022. március 17., helye: https://ec.europa.eu/eurostat/web/products-datasets/-/prc_hicp_midx)

Függelék

Az elemzésem során megírt kódok az alábbi linken érhetők el (a GitHub oldalamon):

https://github.com/Szabad98/MSc_Szakdolgozat_Recovery_Rate_Modelling

Összegző táblázatok wrong way risk modellekhez:

9. táblázat: Összegző statisztikai táblázat wrong-way risk modellhez (Észtország)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Negyedéves valós PD</i>	0,0392	0,0079	0,0219	0,0366	0,0413	0,0440	0,0518
<i>Negyedéves átlagos RR</i>	0,5403	0,1618	0,3503	0,4073	0,4655	0,6970	0,7986
<i>Egy főre jutó GDP (EUR)</i>	4933	626	3860	4522	4875	5302	5302
<i>Munkanélküliségi ráta (%)</i>	5,93	1,15	4,10	5,00	5,90	6,73	7,90
<i>Infláció – HICP (%)</i>	108,06	4,85	99,64	105,05	109,19	110,80	119,60
<i>PD mögöttes faktora</i>	0,0000	2,0711	-3,5157	-1,3888	0,2983	1,3772	5,0006
<i>RR mögöttes faktora</i>	0,0000	0,8771	-1,7123	-0,8220	0,2642	0,7435	1,1305

10. táblázat: Összegző statisztikai táblázat wrong-way risk modellhez (Finnország)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Negyedéves valós PD</i>	0,0845	0,0240	0,0196	0,0783	0,0921	0,0996	0,1109
<i>Negyedéves átlagos RR</i>	0,4637	0,2582	0,1007	0,2665	0,3640	0,7309	0,8561
<i>Egy főre jutó GDP (EUR)</i>	10629	648	9390	10288	10520	11000	12340
<i>Munkanélküliségi ráta (%)</i>	7,86	0,85	6,60	7,03	8,00	8,63	9,20
<i>Infláció – HICP (%)</i>	103,00	1,99	99,80	101,30	103,00	104,00	107,50
<i>PD mögöttes faktora</i>	0,0000	2,3838	-4,5248	-2,1456	0,2804	1,5973	5,3028
<i>RR mögöttes faktora</i>	0,0000	0,4817	-0,7356	-0,3780	-0,0563	0,4322	1,0128

11. táblázat: Összegző statisztikai táblázat táblázat wrong-way risk modellhez (Spanyolország)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Negyedéves valós PD</i>	0,1252	0,0488	0,0410	0,0939	0,1291	0,1543	0,2255
<i>Negyedéves átlagos RR</i>	0,4998	0,3642	0,1000	0,2264	0,2378	0,9597	0,9785
<i>Egy főre jutó GDP (EUR)</i>	6262	368	5310	6042	6305	6490	6920
<i>Munkanélküliségi ráta (%)</i>	16,10	2,00	13,60	14,43	15,60	17,02	20,40
<i>Infláció – HICP (%)</i>	103,30	2,63	98,00	101,60	103,60	104,50	110,20
<i>PD mögöttes faktora</i>	0,0000	2,2786	-4,9606	-1,2497	0,2044	1,6630	4,9210
<i>RR mögöttes faktora</i>	0,0000	0,5530	-1,8746	-0,1182	0,0729	0,3470	0,5992

Összegző táblázatok regressziós modellekhez (bimodalitás vizsgálata):

12. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Teljes adatbázis)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Életkor</i>	41,0	12,7	18,0	31,0	40,0	50,0	71,0
<i>Credit-rating alapú PD</i>	0,2899	0,1455	0,0014	0,1774	0,2831	0,3681	0,9944
<i>EAD (ezer EUR)</i>	2439	2042	2	753	2055	3420	10632
<i>Havi törlesztőrészlet</i>	134,3	121,4	2,6	47,3	113,3	178,2	2368,5
<i>Adósság/jövedelem arány</i>	11,23%	69,82%	0,00%	3,96%	7,46%	12,63%	7504,35%

13. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Észtország)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Életkor</i>	37,5	12,3	18,0	28,0	35,0	46,0	70,0
<i>Credit-rating alapú PD</i>	0,2143	0,0991	0,0014	0,1446	0,2154	0,2810	0,9944
<i>EAD (ezer EUR)</i>	2228	2128	2	526	1514	3153	10632
<i>Havi törlesztőrészlet</i>	101,3	94,3	3,5	24,6	70,7	147,2	1908,5
<i>Adósság/jövedelem arány</i>	10,55%	31,43%	0,00%	3,24%	7,21%	13,84%	4986,00%

14. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Finnország)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Életkor</i>	46,1	12,8	18,0	36,0	47,0	56,0	70,0
<i>Credit-rating alapú PD</i>	0,3185	0,1273	0,0293	0,2186	0,3520	0,3993	0,9200
<i>EAD (ezer EUR)</i>	3294	2192	5	1996	2931	4134	10630
<i>Havi törlesztőrészlet</i>	181,3	130,5	2,6	108,1	148,2	237,8	2368,5
<i>Adósság/jövedelem arány</i>	10,52%	81,69%	0,04%	4,62%	7,32%	11,88%	7504,35%

15. táblázat: Összegző statisztikai táblázat táblázat regressziós modellhez (Spanyolország)

	átlag	szórás	min.	Q1	Q2	Q3	max.
<i>Életkor</i>	40,9	10,9	18,0	33,0	40,0	48,0	71,0
<i>Credit-rating alapú PD</i>	0,3880	0,1622	0,0405	0,2966	0,3548	0,4614	0,9206
<i>EAD (ezer EUR)</i>	1789	1193	3	531	1936	2125	10629
<i>Havi törlesztőrészlet</i>	136,1	133,1	14,9	69,2	113,3	142,0	2277,3
<i>Adósság/jövedelem arány</i>	13,25%	97,54%	0,03%	4,72%	7,86%	11,71%	7058,00%