

Budapesti Corvinus Egyetem

**A gyermekvállalás főbb befolyásoló tényezői Kelet-Európában, 2006–2019
között**

Key Determinants of Childbearing in Eastern Europe, 2006–2019

Dolgozat nyelve: Magyar

Készítette: Banai Gellért

BSc Alkalmazott közgazdaságtan III. évfolyam

2024

Témavezető: Kovács László és Rácz Olivér Miklós

Absztrakt

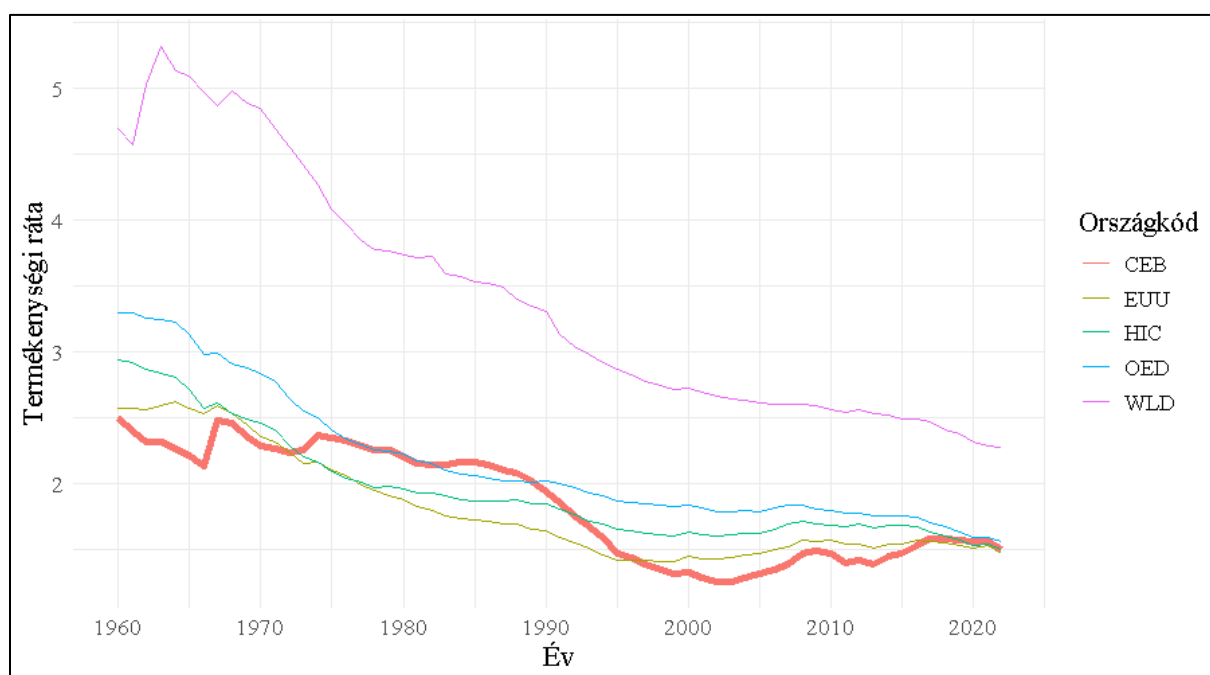
A tartósan alacsony termékenységi ráta egyre növekvő társadalmi és gazdasági kihívás a fejlett országokban. Kutatásomban a különböző 2006-2019 közötti állami beavatkozások (pl. szülési szabadság, családi szociális kiadások), valamint a különböző társadalmi és gazdasági tényezők hatását vizsgáltam a termékenységi rátára bayesi strukturális egyenletmodell (BSEM) segítségével Kelet-Európa 9 EU-s országában. Eredményeim alapján a nők felsőoktatásban való részvétele befolyásolja leginkább a gyermekvállalást, valamint megerősítést nyert, hogy a mélyszegénységben élők kevésbé hajlamosak gyermeket vállalni. Kutatásomnak több limitációja is akad, főképp a kis minta és több fontos változó hiánya (a hiányzó értékek magas aránya miatt) torzíthatja az eredményeket.

Tartalom

I. Bevezetés.....	1
II. Kutatási kérdés	2
III. Szakirodalom.....	3
IV. Adatok	7
V. Módszertan	9
Lasso-regresszió	9
Bayesi strukturális egyenletmodell (BSEM)	11
VI. Eredmények.....	19
Változószelekció.....	19
BSEM modell	20
Limitáció.....	26
VII. Konklúzió	26
Melléklet.....	28
Irodalomjegyzék.....	28

I. Bevezetés

A világ szinte összes országában megfigyelhető, hogy az elmúlt évtizedekben csökkent a termékenységi rátájuk szintje. Ezt jól érzékelteti, hogy míg 1960-ban kb. 5 volt a világ országainak termékenységi rátáinak átlaga, addig ez a szám 60 év elteltével csupán 2,3 körül található és a tendencia továbbra is negatív. Különösen drámai a helyzet Európában és a Távol-Kelet egyes országaiban. Míg 1970-ben 2,8 volt az átlagos termékenységi ráta az OECD országokban - mely szervezet alapvetően magában foglalja az előbb említett régiókat -, addig napjainkban ez a szám 1,5-re esett vissza (ld. 1. ábra), amely jóval a társadalom reprodukciójához szükséges szint alatt helyezkedik el¹. Egy ország népességszáma akkor fenntartható, ha teljes termékenységi arányszáma eléri a 2,1-et. Amennyiben a termékenység tartósan a reprodukciós szint alá kerül, az össznépesség csökken és öregszik – pozitív migrációs egyenleg nélkül. A reprodukciós szinthez közeli tartós termékenység kiegyensúlyozott korösszetételű népességet eredményez, ami nélkülözhetetlen feltétele a szociális ellátórendszerek hosszú távú fenntarthatóságának².



1. ábra: TFR változása a Kelet-Európában, a fejlett országok körében és a világban (1960-2022)

Forrás: saját szerkesztés

¹ Forrás: <https://data.oecd.org/pop/fertility-rates.htm>

² Forrás: <https://ksh.hu/s/kiadvanyok/fenntarthato-fejlodes-indikatorai-2022/1-2>

Egy EU-s közvéleménykutatás³ szerint a népesség elöregedése (42%), valamint a munkaképes korú lakosság csökkenése és a munkaerőhiány (40%) azok a demográfiai kihívások, amelyeket a válaszadók saját tagállamaikban a legégetőbbnek tartanak. Az Unió polgárai egyetértenek abban, hogy az EU jelenlegi demográfiai trendje veszélyezteti a hosszú távú gazdasági jólétet és versenyképességet (69%), valamint hozzájárul a munkaerőhiányhoz (67%). A válaszadók 85%-a egyetért azzal, hogy a demográfiai változások kezelése szoros együttműködést igényel minden érintett kormányzati szint között.

A magas és közepes jövedelmű országok körében a tartósan alacsony termékenységi ráta egyre növekvő kihívást jelent és ezen demográfiai probléma súlyos gazdasági következményekkel is jár (Magnus, 2008). Hosszú távon a társadalom elöregedésének következtében – pozitív migrációs egyenleg nélkül - ezen országokban valószínűleg romlani fog az életszínvonal. Az idősödő társadalmak – ahol növekszik az eltartottak és csökken az eltartók aránya – egyre nagyobb terhet rónak a nyugdíjrendszerekre és az egészségügyi kiadásokra. Széleskörű reformok nélkül ezek a rendszerek egy idő után fenntarthatatlanná válnak. Emellett az idősödő munkaerő csökkenő termelékenységgel és az innoválás csökkenésével járhat, ami a gazdaság lassulásával fenyeget.

Véleményem szerint a XXI. században gyermekvállalási kedv összeomlása egyfajta piaci kudarc, amelyet a piac nem tud jól allokálni és a felnőtt népesség önérdékkövető magatartása nem a kívánt társadalmi optimumhoz vezet. Mindazon jóléti veszteségek, amelyet a születésszám deficit okoz indokolhatják az állami beavatkozást a gyermekvállalás „piacára”, s ezen széleskörű gazdasági következményektől tartván a kormányok szerte Európában igyekeznek is a gyermekvállalást elősegítő politikákat alkalmazni.

II. Kutatási kérdés

A kutatás az állami beavatkozások - közvetett (public services pl. oktatás, és public goods pl. politikai stabilitás, infrastruktúra fejlesztés) és a közvetlen (normatív pénzbeli támogatások, pl. szülési szabadság, szociális ráfordítás) támogatások -, valamint a különböző gazdasági, társadalmi, politikai tényezők hatását vizsgálja a gyermekvállalási hajlandóságra Kelet-Európában, 2006-2019 között. A cél annak meghatározása, hogy mely tényező befolyásolja leginkább a gyermekvállalást, valamint az állami beavatkozásnak van-e hatása a születésszám növelésében. A kutatás hozzájárulhat a gyermekvállalás ösztönzésére irányuló közpolitikai

³ Forrás: <https://europa.eu/eurobarometer/surveys/detail/3112>

döntések megalapozásához, különösen a kelet-európai országok körében (mint például Magyarország). Alkalmazott ökonometriai módszertanom a bayesi strukturális egyenletmodellezés (BSEM).

III. Szakirodalom

Gary Becker 1960-ban született műve előtt a gyermekvállalásra úgy tekintettek, mint ami kívül esik a gazdasági elemzés hatókörén. Becker volt az első egyike, aki közgazdasági elemzés tárgyává tette a gyermekvállalás kérdését, s mint gazdasági döntést értelmezte. Az ő keretrendszerében a gyermek is egyfajta normál jószág (Black et al., 2013), mint például az autók vagy a házak. Az iránta való keresletet is úgy modellezi, mint a normál jószágok esetén: a gyermek „árához” köti. Ez az „ár” lehet közvetlen költség, pénzbeli, mint az iskoláztatás és közvetett, mint például a szabadidő, mely a gyermeknevelés alternatíva költsége. Ezen neoklasszikus modell szerint a gyereknevelés költségeinek csökkenése (az állami támogatás eredményeként) vagy a jövedelmek bármilyen növekedése várhatóan növeli a gyermekek iránti keresletet (Becker 1981; Cigno 1991). Az olyan szakpolitikák, mint a családi pótlék, a gyermekek után járó adókedvezmények, a gyermekgondozási támogatások, valamint a szülési és szülői szabadságra vonatkozó juttatások várhatóan mind pozitív hatást gyakorolnak a termékenységre azáltal, hogy csökkentik a gyermekek közvetlen vagy közvetett alternatívaköltségeit, illetve növelik az egyének jövedelmét. Ezen gazdasági modell nagy hatással volt a szakirodalomra, s a szakpolitikák és demográfia között feltételezett kapcsolat középpontjában áll (Gauthier, 2007). Becker szerint van egy trade-off a gyermekek „mennyisége” és „minősége” között, s a szülők nemcsak a gyermekek számát, hanem a minőségét is figyelembe veszik. Ez magyarázhatja azt, hogy a pénzbeli támogatásoknak miért van minimális hatása a fertilitásra. Több kritika is érte Beckert, hogy a megközelítése figyelmen kívül hagyja a biológiai, érzelmi és társadalmi tényezőket (Lee, 2015; Doepke, 2015). Emellett azzal a furcsa összefüggéssel szembesült Becker, hogy a jövedelem növekedése épphogy nem magasabb termékenységhez vezet, hanem kevesebb gyereket vállalnak a párok, amit gazdasági modellje nem magyaráz meg.

A termékenység gazdasági modelljének második feltevése az, hogy az egyének teljes információval rendelkeznek a gyerekvállalás előnyeiről és költségeiről. Ezt a feltételezést számos kutató megkérdőjelezte: valószínűbb a tökéletlen információs helyzet (Goldthorpe 2000).

Úgy vélem, hogy a gyermekvállalás egy nagyon komplex döntés, aminek biztosan van gazdasági aspektusa, de a különböző társadalmi, politikai folyamatok, valamint az ember reprodukcióhoz fűződő vágya is magyarázhat.

Zhang és társai (2023) empirikus tanulmányukban az OECD országainak családi jóléti politikáinak termékenységsztönző hatását vizsgálták regressziós elemzés, szürke relációs analízis (GRA) és a fuzzy halmaz kvalitatív komparatív elemzés (fsQCA) módszereivel. Az eredmények szerint:

- (1) A családi jóléti politikák jelentősen növelik a termékenységet, és ez a növekedési hatás hosszú távon is fennmarad. Azonban ez a növekedés gyengül azokban az országokban, ahol a termékenységi ráták 1,5 alatt maradnak.
- (2) A jóléti politikai intézkedések hozzájárulása a termékenységsztönző hatáshoz országonként eltérő. A készpénzes juttatások hozzájárulása a legmagasabb a világ országainak több mint felében, a releváns szolgáltatások és természetbeni kiadások hozzájárulása a legmagasabb az országok 29%-ában, míg az adókedvezményes kiadások hozzájárulása a legmagasabb az országok 14%-ában.
- (3) Emellett kiemelik, hogy a foglalkoztatás a fő eszköz a családi jövedelem biztosítására, s kritikus jelentőséggel bír a család fenntartásában, ezért elengedhetetlen a fiatalok munkanélküliségének csökkentése és a fiatalok foglalkoztatási minőségének javítása.

A Hétfa Intézet kutatása (2019) szerint Magyarországon a termékenységi döntéseket elsősorban a foglalkoztatás, a jövedelem és a lakhatási kilátások befolyásolják (Szabó-Morvai, Balás, Bördös, & Herczeg, 2019). Emellett a nők jelenlegi foglalkoztatottsága, az óvoda elérhetősége és a rugalmas munkalehetőségek is jelentősen növelik a születési valószínűséget. A családi adókedvezménynek köszönhető jövedelem növekedés, valamint az otthonteremtési támogatások is pozitív hatással vannak a termékenységre.

Mivel a családtámogatások csökkentik a gyermekvállalás költségeit, elvárható, hogy a magasabb támogatások növeljék a termékenységet. Blanchet és Ekert-Jaffé (1994) tanulmányában bemutatott empirikus bizonyítékok részben alátámasztják ezt a hipotézist. Regressziós eredményeik szerint körülbelül 0,04 gyermekkel nőne a termékenység nőnként, ha a családi juttatások globális indexe 25 százalékkal nőne. Az ő tisztán statikus modelljükben nincs különbségtétel a rövid és hosszú távú hatások között.

Hasonló módszertant alkalmazva, s az országok körét kibővítve Anne Gauthier és Jan Hatzius (1997) 1970-1990 közötti időszakot felölelő, 22 iparosodott ország adatait alapul véve, GMM modelljük szerint a pénzbeli támogatások pozitívan függnak össze a termékenységgel. A hatás azonban erősen korlátozott. A családi támogatások 25 százalékos növekedése körülbelül 0,6 százalékkal magasabb termékenységi szintet eredményezne rövid távon, és körülbelül 4 százalékkal magasabbat hosszú távon, tehát átlagosan körülbelül 0,07 gyermekkel nőne nőnként a termékenység. Ezt a hatást országonként változónak találták. Az eredményeik arra utalnak, hogy a támogatások rövid távú hatása kisebb, míg a hosszú távú hatás nagyobb, mint amit Blanchet és Ekert-Jaffé becsült. Azonban a szülési szabadság (mind az időtartam, mind a támogatás) nem volt szignifikáns tényező a termékenység alakulásában.

A releváns tanulmányok, amelyek idősoros adatokat használnak (Ermisch, 1988; Zhang et al., 1994), vagy összevont (pooled) adatokat (Ekert, 1986; Caudill & Mixon, 1993; Blanchet & Ekert-Jaffé, 1994), általában pozitív kapcsolatot találnak a termékenység és a családpolitikák között.

A makroszintű adatokat használó tanulmányok arra jutottak, hogy a pénzbeli juttatások termékenységre gyakorolt hatása inkább a születések időzítésére, mintsem a gyermekek teljes számára van hatással. Ermisch (1988) az életkor- és paritás-specifikus termékenységi rátákat vizsgálva azt találta, hogy a gyermeknevelési támogatások emelkedése Nagy-Britanniában növelték a magasabb paritású születések valószínűségét, de egyben ösztönözték a fiatal anyaságot is. A szakpolitikák időzítési hatását figyelte meg Svédországban Hoem (2005) és Andersson és társai (2006) is a szülői szabadságra vonatkozó támogatással kapcsolatban. Björklund (2006) ugyancsak svéd adatokon megfigyelte, hogy a támogatások kiterjesztése növelte a termékenység szintjét, lerövidítette a születések közötti időbeli távolságát, és ingadozásokat idézett elő a termékenységi rátákban, ugyanakkor nem változtatta meg a nők iskolai végzettsége és a befejezett termékenység közötti negatív kapcsolatot.

A mikroszintű adatokon alapuló tanulmányok eredményei többször ellentmondásosak: alapvetően pozitív kapcsolatot feltételeznek a pénzbeli támogatások és a termékenység között (Ekert-Jaffé et al., 2002), de több tanulmány is rámutat, hogy az alacsonyabb családtámogatási rendszerű országokban is sokszor hasonló a gyermekek születési aránya (Breton & Prioux, 2005)

A munkával kapcsolatos szakpolitikáknak - mint például a szülési és szülői szabadságot vagy a gyermekgondozási lehetőségeket - eredményei vegyesek: egyes kutatások arra a következtetésre

jutottak, hogy a kis mértékben pozitívan befolyásolják a termékenységet, míg mások nem találtak bizonyítékot arra, hogy ezen szakpolitikák hatással lennének a termékenységre.

A szülési és szülői szabadsággal kapcsolatban Hyatt és Milne (1991) kanadai adatok alapján történő becslése szerint a szülési támogatás reálértékének 1%-os növelése a teljes termékenységi rátát 0,09% és 0,26% közötti mértékben növelheti. Rønsen (2004) finn és norvég adatokon, valamint Hoem (1993) svéd adatokon végzett kutatásai szintén arra a következtetésre jutottak, hogy a szülői szabadság pozitív hatással van a termékenységre. Ezzel szemben Hoem és társai (2001) osztrák adatokon végzett tanulmánya nem talált hatást a szülői szabadság változásainak termékenységre gyakorolt hatására, kivéve a harmadik gyermek megszületésének időzítését, amelyet előrébb hozott.

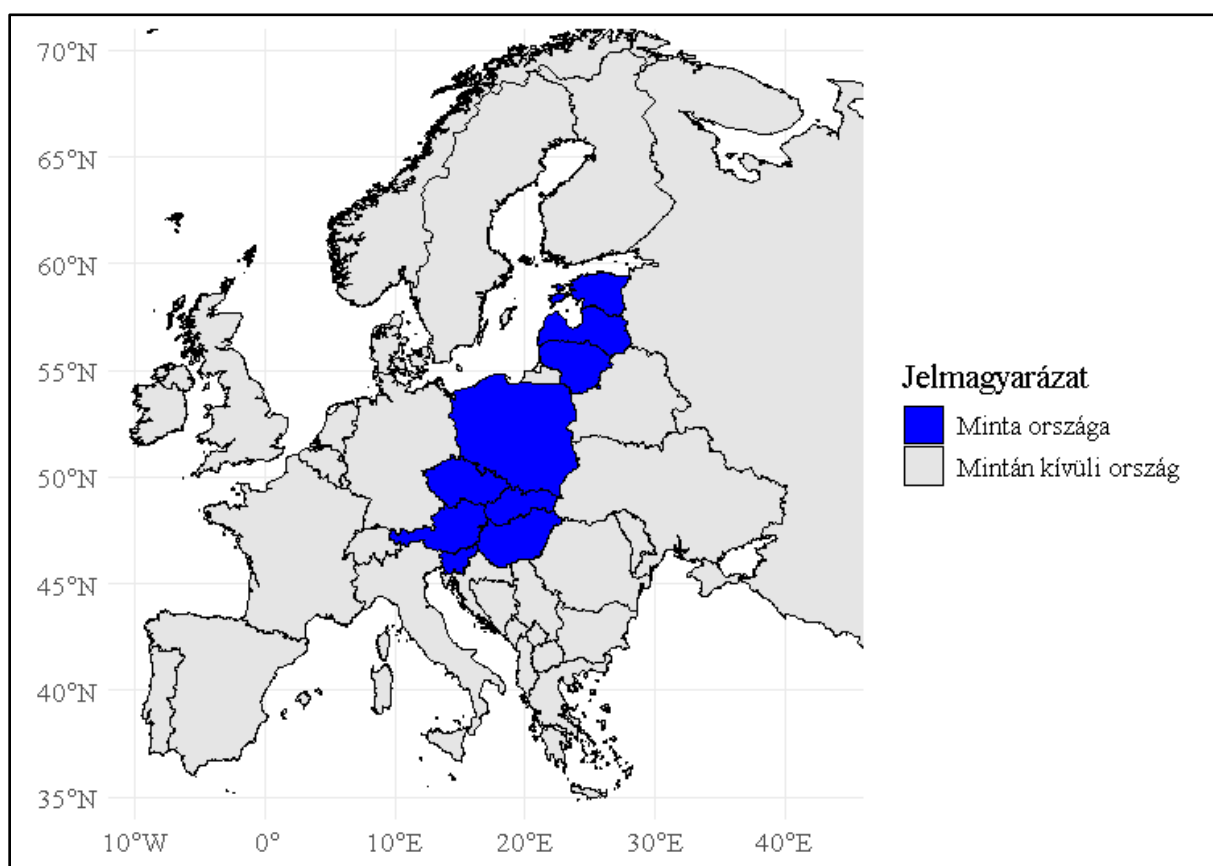
A gyermekgondozás elérhetőségének és költségének tekintetében szintén vegyes eredmények születtek. Norvég makroszintű adatok alapján Kravdal (1996) becslése szerint a gyermekgondozási ellátás 20 százalékpontos növekedése legfeljebb 0,05 gyermekkel növelheti a teljes kohorsztermékenységet egy nőre vetítve. A gyermekgondozási költségek csökkenésének és a gyermekgondozási lehetőségek bővülésének pozitív hatását figyelték meg Pacalova és társai (2003), valamint Del Boca és társai (2003) több országot vizsgáló elemzéseikben. Hank és Kreyenfeld (2003) eredményei alapján az informális gondozási formákhoz való hozzáférés növeli a szülővé válás valószínűségét.

Ugyanakkor Rønsen (2004) Norvégiára és Finnországra, Hank és Kreyenfeld (2003) Németországra, valamint Andersson és társai (2004) Svédországra vonatkozóan nem találtak statisztikailag szignifikáns hatást a gyermekgondozási jellemzők (költségek és elérhetőség) és a termékenység között.

Azonban a vágyott és befejezett termékenység közötti rés Európában ma is átlagosan 0,3-0,4 gyermek nőnként (Berghammer & Beaujouan, 2019). Chesnais (1996) szerint „az ideális és a valóság közötti szakadék (a gyermekek száma tekintetében) azt mutatja, hogy a közpolitikák nem tudták elhárítani a termékenységi vágyak megvalósulása előtt álló akadályokat”. Értelmezésem szerint tehát van még hely a kormányzati beavatkozásoknak, így releváns és kiemelkedően fontos területe ez a közgazdaságtannak.

IV. Adatok

A kérdés tanulmányozásához egy egyedi adatbázist építettem, amely több adatforráson alapul: az Eurostaton⁴, a Világbankon⁵ és az OECD⁶ adatbázisán. Panel adatbázisomban összesen 126 megfigyelést vizsgálok, mely 9 EU-s adatait tartalmazza - 2006-2019 közötti időszakban, éves bontásban. Ezen országok: Ausztria, Csehország, Észtország, Magyarország, Lettország, Litvánia, Lengyelország, Szlovákia, Szlovénia. (lsd. 2. ábra). A hiányzó adatok jelentős aránya miatt sajnos nem tudtam több országot is bevonni az elemzésbe (pl. Románia, Bulgária...stb.).



2. ábra: Minta országai térképen

Forrás: saját szerkesztés

Az elemzés függő változója a teljes termékenységi arányszám, amely azt mutatja meg, hogy ha az adott év termékenységi adatai állandósulnának, akkor egy nő élete folyamán átlagosan hány

⁴ Forrás: <https://ec.europa.eu/eurostat>

⁵ Forrás: <https://www.worldbank.org/en/home>

⁶ Forrás: <https://www.oecd.org/en/data/datasets/social-expenditure-database-socx.html>

gyermeknek adna életet⁷. Egy hipotetikus generáció befejezett termékenységet jelzi, amelyet az adott évi női korszpecifikus termékenységi arányszámok összegzésével számítottunk ki.⁸

A szakirodalom alapján a lentebbi tényezők hatnak leginkább a gyermekvállalásra, amelyeket az alábbi változókkal igyekszek megragadni:

Tényező	Változó	Hiányzó értékek aránya	Forrás
Nők helyzete	Nők munkaerőpiaci részvétele (15+ életkorú nők %-ában)	-	Világbank
	Felsőoktatásban részesült nők aránya	-	Eurostat
	Nők életkora az első gyermek születésekor	-	Eurostat
	Első házasságok aránya	4%	Eurostat
Életszínvonal	Egy főre jutó GDP, PPP (constant 2021 international \$)	-	Világbank
	Nem várt kiadásokra való képtelenség (társadalom %-a)	-	Eurostat
	Szegénységből meneküléshez szükséges órák száma	-	Eurostat
	Várható élettartam (év)	-	Világbank
Gazdasági ciklus	HCPI index éves változása	-	Eurostat
	Fiatalok foglalkoztatási rátája (%)	-	Eurostat
	Munkanélküliség (%)	-	Világbank
	Egy főre jutó GDP növekedés (évi %)	-	Világbank
Közzolgáltatások és közjavak	Egészségügyi kiadások (GDP %-ában)	-	Világbank
	Kormányzati költség közművekre és lakhatásra (USD 2015 PPP/fő)	-	OECD
	Kormányzati költség rekreációra, kultúrára és vallásra (USD 2015 PPP/fő)	-	OECD
	Politikai stabilitás index	-	Világbank
Népesség	Népesség a legnagyobb városban (városi népesség %-ában)	-	Világbank
	Népsűrűség (fő/km ²)	-	Világbank
	Népességnövekedés (évi %)	-	Világbank
Szülés	Gyermekehalandóság (1,000 élve születésre)	-	Világbank
	Anyasági halálozás kockázata (%)	-	Világbank
Ingatlanpiac	Lakásárindex	9%	Eurostat
	Jövedelemből az albérletre költött arány (%)	2%	Eurostat
Szabadidő	Minimum szabadság	-	Wikipédia
	Hivatalos ünnepnapok száma	-	Wikipédia
	Fizetett apaszabadság	-	OECD
	Fizetett szülési szabadság	-	OECD
Támogatások	Kormányzati költség lakhatási támogatásra (USD 2015 PPP/fő)	-	OECD
	GMI	-	OECD
	Családokra fordított szociális védelmi kiadások (GDP %-ában)	-	Eurostat
	Adópolitika hatás	1%	OECD.

1. táblázat: Magyarázóváltozók

Forrás: saját szerkesztés

⁷ Forrás: <https://demografia.hu/hu/tudastar/fogalomtar/38-teljes-termekenysegi-aranszam>

⁸ Forrás: <https://ksh.hu/s/kiadvanyok/fenntarthato-fejlodes-indikatorai-2022/1-2>

A független változók közül mind a 31 kvantitatív változó. Elemzésem egyik legfontosabb limitációja, hogy néhány fontos változó esetében nincs elérhető adat (pl. gyermekgondozásra töltött idő, társadalmi normák változása, vallásosság).

Emellett néhány magyarázóváltozó esetében fellépnek hiányzó értékek⁹, amelyeket láncolt egyenletek ('Multivariate Imputation by Chained Equations') segítségével¹⁰ pótoltam, ezzel csökkentve az adathiány okozta torzítást, s növelve az elemzés pontosságát. E módszer lényege, hogy iteratív módon, a láncolt egyenletek módszerét használva - ami azt jelenti, hogy minden hiányzó értéket többszörös imputációval helyettesít a többi változó alapján - tölti ki a hiányzó értékeket különböző modellek (pl. Predictive Mean Matching (PMM), logisztikus regresszió) segítségével. Esetemben a PMM módszert használtam imputálásra. A hiányzó értékek aránya az összes megfigyelésen belül 1% és 9% közötti intervallumon alakul, azonban csak olyan változók kerültek be a végső modellbe, ahol nincs hiányzó érték.

V. Módszertan

Lasso-regresszió

Első alkalmazott módszertanom a Lasso-regresszió, melyet változószelekcióra használtam. E módszert Robert Tibshirani (1996) statisztikus fejlesztette és publikálta először, mely az első regularizációs együtthatóbecslési eljárás, amely változószelekcióra is képes (Kovács, 2021). A Lasso (Least Absolute Shrinkage and Selection Operator) egy gyakran alkalmazott eljárás a gépi tanulásban, különösen magas dimenzionalitású adatok esetén, mivel automatikus változószelekciót tesz lehetővé, és a becslés hatásos olyan esetekben is, amikor a paraméterek száma meghaladja a megfigyelések számát¹¹. A módszer lényege, hogy a regressziók legkisebb négyzetek elvű becslése során a reziduális négyzetösszeghez (Residual Sum of Squares, RSS) egy büntetőtagot ad hozzá, amelyet egy regularizációs hiperparaméterrel (λ , $\lambda \in (0,1)$) szoroznak meg. Ez a paraméter szabályozza a regularizáció mértékét: minél nagyobb a λ értéke, annál erősebben érvényesül a büntetés, így a regressziós együtthatók egyre inkább a nullához közelítenek. Ennek eredményeként egyes változók jelentősége csökken vagy teljesen kiesnek a modellből (együtthatójuk becsült értéke pontosan 0) – ez automatikus

⁹ Ezen változók: Nők felsőfokú oktatása, Népsűrűség, Nagy városokban élő aránya, Nők munkaerőpiaci részvétele, Szülési halál kockázata, HCIP, Szociális védelmi kiadások, Nem várt kiadásokra való képtelenség

¹⁰ R programozási nyelv MICE csomagja

¹¹ Forrás: <https://www.ibm.com/think/topics/lasso-regression>

változószelekcióhoz vezet. Ezzel szemben alacsony lambda érték esetén a büntető hatás gyengül, így több változó maradhat a modellben. Ez a büntetési mechanizmus elősegíti a modell ritkaságát (sparsity), amely segít elkerülni a túltanulást (overfitting).

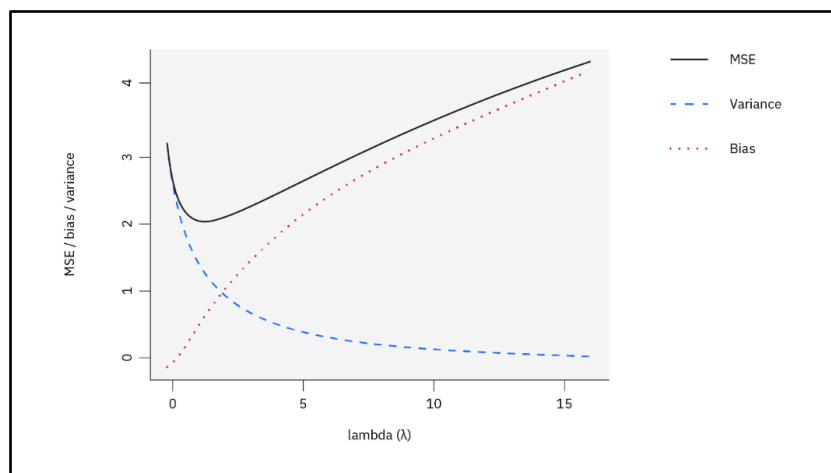
$$\min_{\beta} ||y - X\beta||^2 + \lambda \sum_{j=1}^p |\beta_j|$$

1. egyenlet: Lasso

Forrás: saját szerkesztés

A bias-variance tradeoff (torzítás és variancia közötti „kompromisszum”) a prediktív modellek jól ismert jellemzője. A torzítás azt méri, hogy a modell becslései átlagosan mennyire térnek el a valós értékektől, míg a variancia azt mutatja meg, hogy egy adott modell különböző mintákon mennyire eltérő becsléseket ad. A magas torzítás alulillesztéshez (underfitting), míg a magas variancia túlillesztéshez (overfitting) vezet. E két mutató egyidejű csökkentése nem mindig lehetséges – ezért van szükség regularizációs technikákra, mint a lasso-regresszió.

A Lasso-regresszióban a lambda (λ) hiperparaméter – más néven L1-büntetés – biztosítja a torzítás és variancia közti egyensúlyt. Amint λ nő, a torzítás is nő, a variancia viszont csökken, így egy egyszerűbb, kevesebb paramétert tartalmazó modell jön létre. Ezzel szemben csökkenő λ esetén a modell komplexebbé válik, több paraméterrel. Ha λ értéke nulla, akkor az eredmény megegyezik az OLS lineáris regressziós modellel (regularizáció nélkül).



3. ábra: Torzítás-variancia tradeoff

Forrás: <https://www.ibm.com/think/topics/lasso-regression>

A Lasso-regresszió célja az átlagos négyzetes becslési hiba, azaz MSE^{12} minimalizálása úgy, hogy közben kiegyensúlyozza a torzítás és a variancia közötti ellentétet. Ezt úgy éri el, hogy a maradéknégyzetek összegéhez hozzáadja az együtthatók abszolút értékeinek összegét, megszorozva λ -val. A lambda hiperparaméter kiválasztására számos módszert dolgoztak ki, azonban a k-fold keresztvalidáció vitathatatlanul a legelterjedtebb (Arlot & Celisse, 2010), így én is ezt használom.

A Lasso különösen jó akkor, ha a prediktorváltozók száma nagyobb, mint a megfigyelések száma, így segíthet a dimenzionalitás csökkentésében azáltal, hogy nullára zsugorít bizonyos súlyokat, így kizárja a kevésbé fontos változókat. Az L1-büntetés által bevezetett torzítás mesterségesen nullához húzza az együtthatókat. Egyes változók együtthatója nullává válik, így a modell csak a legfontosabb prediktorokat tartja meg. Azonban, fontos megjegyezni, hogy Zhao - Yu (2006) állítása szerint a Lasso-módszer nem megfelelően kezeli a multikollinearitás jelenségének káros hatásait a modellszelekció során.

A Lasso-regresszió által szelektált változók hatását a függő változóra egy bayesi strukturális egyenlet modellben (BSEM) becsültem meg.

Bayesi strukturális egyenletmodell (BSEM)

A Bayes-féle módszerek viszonylag későn jelentek meg a strukturális egyenletmodellezés (SEM) területén, a huszadik század végén és a huszonegyedik század elején (Kontakos & Tsairidis, 2024). A nagyteljesítményű számítógépes és algoritmikus eszközök, valamint olyan statisztikai csomagok fejlesztése, amelyek képesek kezelni a posterior eloszlásokból származó szimulációs modelleket, lehetővé tették a Bayes-féle módszerek alkalmazhatóságát. Több kutatás kimutatta, hogy a Bayes-féle strukturális egyenletmodellek jobb becslőket adnak, mint a klasszikus elmélet alkalmazása (Lee & Song, 2004; Jackman, 2000). Press (2003) tanulmányában arra a következtetésre jutott, hogy a Bayes-féle strukturális egyenletmodellek jobban teljesítenek olyan esetekben, amikor a klasszikus SEM nem működik jól, például kis minták esetén, vagy amikor nemlineáris kapcsolatokat kell feltételezni a látens változók között. Az én esetemben is kis mintával ($N=126$) dolgozok, ezért esett a választásom a klasszikus SEM helyett a bayesi SEM-re.

A BSEM esetén Y egy megfigyelt valószínűségi változó, amelynek megfigyelt értékeit y vektorban tároljuk. Továbbá, θ egy ismeretlen paraméter vagy paramétervektor, amely a

¹² Az MSE az előre jelzett és a valós függő változók értékei közötti különbséget méri.

vizsgált populációhoz tartozik. A cél az ismeretlen paraméter becslése az összegyűjtött adathalmaz, azaz az Y valószínűségi változó y megfigyelt értékei alapján. Ez a becslés a paraméter valószínűségi eloszlása alapján történik az adatok ismeretében, amelynek jelölése: $L(\theta | y)$. Itt rejlik a különbség a klasszikus módszer és a Bayes-féle megközelítés között. A klasszikus módszer szerint a θ paraméter ismeretlen, de konstans értéke van, amit a kutató kíván meghatározni. Ezzel szemben a Bayes-féle módszer a θ paramétert valószínűségi változónak tekinti, amely eloszlással rendelkezik, és ez az eloszlás tükrözi a paraméterrel kapcsolatos bizonytalanságot. Így mind az adatok megfigyelt értékei, mind a paraméter valószínűségi változók és együttes valószínűségi sűrűségfüggvényük: $p(\theta, y)$.

A feltételes valószínűség definíciója alapján:

$$p(\theta, y) = p(y|\theta) \cdot p(\theta)$$

2. egyenlet: feltételes valószínűség

És mivel:

$$p(y | \theta) \cdot p(\theta) = p(\theta | y) \cdot p(y)$$

Ebből levezethető a Bayes-tétel:

$$p(\theta|y) = \frac{p(\theta|y)}{p(y)} = \frac{p(y | \theta) \cdot p(\theta)}{p(y)}$$

3. egyenlet: Bayes-tétel

A $p(\theta | y)$ feltételes valószínűség az ismeretlen paraméter posterior eloszlása az adatok y ismeretében. A $p(\theta)$ az ismeretlen paraméter prior eloszlás, $p(y | \theta)$ a minta likelihoodja és $p(y)$ a minta tapasztalati eloszlása. Mivel a nevező $p(y)$ nem tartalmazza az ismeretlen paramétert, s az eloszlás, ha y rögzített, akkor a θ -ra nézve a következő kapcsolat adódik, amely a Bayes-módszertan alapját képezi:

$$p(\theta | y) \propto L(\theta | y) \cdot p(\theta)$$

A fő különbség a klasszikus modellek és a bayesi modellek között az, hogy az utóbbi képes kezelni a becsülendő paraméterek eloszlásának priorját. A probléma az, hogy ezeket az eloszlásokat úgy kell meghatározni, hogy a lehető legjobb becslést adjuk a paraméterekre. Kétféle ilyen eloszlás létezik. Az informált priorok és a nem-informált priorok. Az informált prior eloszlások azok, amelyeket a kutató akkor használ, amikor világos vagy elegendő jelzés áll rendelkezésre a becsülendő paraméterek eloszlásáról. Ilyen világos jelzéseknek tekinthetők például a paramétereloszlás alakjáról vagy formájáról szóló előzetes ismeretek, a korábbi, hasonló adatokon végzett kutatások alapján. Az informált eloszlás egy speciális típusa a konjugált prior eloszlás. Azt a tulajdonságot, amely szerint az utólagos (posterior) eloszlás ugyanolyan paraméteres alakban megadható, mint a prior eloszlás, konjugációnak nevezzük, és a prior eloszlást konjugált prior eloszlásnak hívjuk. Általában az informált eloszlások saját paraméterekkel rendelkeznek, ezeket hiperparamétereknek nevezzük (Lee, 2007).

Ellentétben az informált prior eloszlásokkal, amikor a kutatónak nincs elegendő vagy világos tudása/információja a becsülendő modell paramétereiről, akkor az általa használt prior eloszlásokat nem-informált eloszlásoknak nevezzük. Az ilyen eloszlások meglehetősen nagy szórással rendelkezhetnek, vagy lehetnek állandóak (pl. egyenletes eloszlás). Ilyen esetekben a populáció paramétereinek becslését nem befolyásolja semmilyen előzetes (azaz a priori) információ, a 3. egyenlettel adott posterior eloszlás meghatározásában a minta likelihoodja dominál. Ezért a nem-informált eloszlások használata hasonló eredményeket ad, mint a klasszikus maximum likelihood módszertan (Lee, 2007).

A strukturális egyenletmodellezésben (SEM) a becsülendő ismeretlen paraméter egy vektor, jelöljük θ -vel, amely a becsülendő modell egyedi paramétereiből áll, azaz faktor loadingokból, szórásokból, varianciákból és korrelációkból. A valószínűségi integrálok gyakran nem oldhatók meg analitikusan, így a becsléshez numerikus szimulációs módszerekre van szükség, különösen a Markov-lánc Monte Carlo (MCMC) technikákra. Az általános elképzelés a következő: ahelyett, hogy analitikusan megoldanánk és meghatároznánk az utólagos eloszlás momentumait vagy más mennyiségeit a θ paraméter becslése érdekében, az MCMC algoritmusok kifejezetten előállított mintákat vesznek az utólagos eloszlásból a modellparaméterekre a minta likelihoodjának, empirikus eloszlásának és prior eloszlásának szimulációján és a szimulált értékekre a Bayes-tétel alkalmazásával (Gilks, Richardson, & Spiegelhalter, 1996). A bayesi strukturális egyenletmodellezésben ez a technika azt célozza, hogy utólagos eloszlásokat állítson elő az ismeretlen paraméterekre a nem megfigyelhető vagy látens változók esetében a kifejlesztett modellben. Az MCMC technika szinte bármilyen esetben alkalmazható, még akkor

is, ha a célzott eloszlások nagyon összetettek, ami lehetetlenné teszi a hagyományos szimulációs módszerek alkalmazását. A fő különbség az MCMC módszerek és a klasszikus módszerek között az, hogy az MCMC módszer egy függő értéksorozatot (Markov-láncot) állít elő -, ahol minden új érték csak az előzőtől függ - és nem független választások sorozatát (Link & Eaton, 2012).

Az MCMC módszerek egyike a Gibbs-mintavételezés (Geman & Geman, 1984). Legyen $\theta=(\theta_1,\theta_2,\dots,\theta_n)^T$ a modell paramétereinek vektora, amelyet prior sűrűségfüggvény ír le: $p(\theta)$. Továbbá legyen $\theta^{(0)}=(\theta_1^{(0)}, \theta_2^{(0)},\dots, \theta_n^{(0)})^T$ a megfelelő kezdeti értékek vektora. E kezdeti értékvektor alapján a Gibbs-algoritmus retrospektíven előállítja az $\theta^{(s)}$ vektort az $\theta^{(s-1)}$ alapján a következő módon:

$\theta_1^{(s)}$ az $p(\theta_1 | \theta_2^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_n^{(s-1)}, y)$ eloszlásból

$\theta_2^{(s)}$ az $p(\theta_2 | \theta_1^{(s-1)}, \theta_3^{(s-1)}, \dots, \theta_n^{(s-1)}, y)$ eloszlásból

$\vdots$

$\theta_n^{(s)}$ az $p(\theta_n | \theta_1^{(s)}, \theta_2^{(s)}, \dots, \theta_{n-1}^{(s)}, y)$ eloszlásból

ahol $s = 1,2,\dots,S$ a Monte Carlo iterációk száma. A legtöbb feltételes eloszlás, amely megjelenik, standard normális eloszlás vagy Wishart-eloszlás. A következő lépés egy független vektorsorozat létrehozása:

$\theta^{(1)} = \{\theta_1^{(1)}, \dots, \theta_n^{(1)}\}$

$\theta^{(2)} = \{\theta_1^{(2)}, \dots, \theta_n^{(2)}\}$

$\vdots$

$\theta^{(s)} = \{\theta_1^{(s)}, \dots, \theta_n^{(s)}\}$

amely a jól ismert markovi tulajdonsággal rendelkezik, mivel az $\theta^{(s)}$ vektor független a $\theta_1^{(0)}, \dots, \theta_n^{(s-2)}$ értékektől, ha ismert a $\theta^{(s-1)}$. Amint az eljárásból látható, a kulcselemek az alkalmazandó láncok száma és az iterációk száma. Ezek a számok a konvergenciát az eloszlásra vonatkozóan befolyásolják.

A fenti eljárást általános strukturális egyenletmodellekre (SEM) specializálva (Jöreskog & Sörbom, 1996): legyen $\mathbf{y} = (y_1, y_2, \dots, y_p)^T$ a vizsgált modellben szereplő megfigyelt változók $p \times 1$ vektora, és $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_q)^T$ a $q \times 1$ látens változók vektora. Az egyszerűség kedvéért feltételezzük, hogy a $\boldsymbol{\omega}$ vektor tartalmazza mind az exogén látens változókat (ξ), mind az

endogén látens változókat (η). Az exogén és endogén szerkezetek metrikus modelljét a következő egyenlet adja meg:

$$y = \mu + \Lambda\omega + \varepsilon,$$

4. egyenlet: Mérési modell egyenlete

ahol y a megfigyelt változók vektora, μ a metrikus modellben szereplő $p \times 1$ konstansokat tartalmazó vektor, Λ a $p \times q$ terhelési (loading) mátrix, ω a latens változók vektora, és ε a $p \times 1$ véletlen hibák vektora, amely normál eloszlást követ: $N(0, \Psi\varepsilon)$, ahol $\Psi\varepsilon$ egy diagonális kovariancia mátrix.

A metrikus modell strukturális modellje a következő egyenlettel adható meg:

$$\eta = \alpha + B\eta + \Gamma\xi + \delta,$$

5. egyenlet: Strukturális modell egyenlete

ahol a $q_1 \times 1$ méretű η vektor a látens változók endogén részét képviseli, míg a $q_2 \times 1$ méretű ξ vektor a q_2 exogén látens változót ($q_1 + q_2 = q$) tartalmazza. A $q_1 \times q_1$ méretű B mátrix az η vektorban szereplő endogén változók közötti kapcsolatokat írja le. A $q_1 \times q_2$ mátrix Γ az exogén látens változók (ξ) hatását írja le az endogén látens változókra (η). A ξ és az η az ω -t adja vissza. α a strukturális modell $q_1 \times 1$ méretű konstansait tartalmazó vektor, míg δ a strukturális modell hibatagját tartalmazó $q_1 \times 1$ vektor, amely a normál eloszlást követi: $N(0, \Psi\delta)$, ahol $\Psi\delta$ egy diagonális kovariancia mátrix. Továbbá feltételezzük, hogy $\xi \sim N(0, \Phi)$, az $I - B$ mátrix reguláris (nem szinguláris), és hogy a ε és δ hibák egymástól függetlenek.

A modell (4 és 5. egyenlet) előzetes eloszlásainak meghatározásához konjugált prior eloszlást használunk. Legyen Λ_k a terhelési mátrix Λ k -adik sora, és ψ_{ε_k} a $\Psi\varepsilon$ mátrix k -adik diagonális eleme.

Kimutatható (Lee, 2007), hogy a $(\Lambda_k, \psi_{\varepsilon_k})$ és μ konjugált prior eloszlásai:

$$\psi_{\varepsilon_k} \sim \text{IG}(\alpha_{\varepsilon_k}, \beta_{\varepsilon_k}) \text{ (vagy ekvivalensen } \psi_{\varepsilon_k}^{-1} \sim G(\alpha_{\varepsilon_k}, \beta_{\varepsilon_k})),$$

$$\mu \sim N(\mu_0, \Sigma),$$

ahol az $IG(\alpha\epsilon_k, \beta\epsilon_k)$ az Invertált Gamma eloszlás (paraméterei pozitívak). Továbbá:

$$\Lambda_k | \psi\epsilon_k \sim N(\Lambda_k, \psi\epsilon_k H_{\gamma k}),$$

ahol Σ és H pozitív definit mátrixok.

Hasonlóképpen, ha $\psi\delta_k$ a $\Psi\delta$ mátrix k -adik diagonális elemét jelöli, akkor a Φ^{**} és $(\Lambda_k, \psi\delta_k)$ konjugált prior eloszlásai:

$$\Phi \sim IW_{q_2}(R^{-1}, \rho_0) \text{ (vagy ekvivalensen } \Phi^{-1} \sim W_{q_2}(R^{-1}, \rho_0)),$$

$$\psi\delta_k \sim IG(\alpha\delta_k, \beta\delta_k) \text{ (vagy } \psi\delta_k^{-1} \sim G(\alpha\delta_k, \beta\delta_k)),$$

$$\Lambda_k | \psi\delta_k \sim N(\Lambda_k, \psi\delta_k H_k).$$

Ha egyszerűség kedvéért a következő módon jelöljük:

$$\theta_{\text{norm}} = \{\mu, \alpha, \Lambda, B, \Gamma\},$$

(a normál eloszlású paramétervektor)

$$\theta_{IW} = \{\Psi_\epsilon, \Psi_\delta\},$$

(az inverz Wishart eloszlású varianciák és kovarianciák)

akkor:

$$\theta_{\text{norm}} \sim N(\mu, \Omega),$$

$$\theta_{IW} \sim (R, \rho_0).$$

Továbbá Lee (2007) szerint a strukturális egyenletmodellekben (SEM) a latens változók (ω) hiányzó értékeknek tekinthetők. Így a modellekre (4. és 5. egyenlet) vonatkozó minta generálása Gibbs-módszerrel a következő lépésekből áll (Kaplan & Depaoli, 2012):

- | | | | | |
|------|--|-------------|----|--------------|
| I. | $\eta^{(s+1)}$ | mintavétele | az | eloszlásból: |
| | $p(\eta \theta_{\text{norm}}^{(s)}, \theta_{IW}^{(s)}, y)$ | | | |
| II. | $\theta_{\text{norm}}^{(s+1)}$ | mintavétele | az | eloszlásból: |
| | $p(\theta_{\text{norm}} \theta_{IW}^{(s)}, \eta^{(s+1)}, y)$ | | | |
| III. | $\theta_{IW}^{(s+1)}$ | mintavétele | az | eloszlásból: |
| | $p(\theta_{IW} \theta_{\text{norm}}^{(s+1)}, \eta^{(s+1)}, y)$ | | | |

A megfigyelt és látens változók ismeretében a következő lépés a mérési és a strukturális modell paraméterbecslőinek létrehozása az (1) és (2) egyenletek alapján.

PSR

A fent leírt iteratív eljárás konvergenciája ellenőrizhető a Gelman–Rubin-féle konvergenciakritérium segítségével, a PSR mutató alkalmazásával (Gelman & Rubin, Inference from Iterative Simulation Using Multiple Sequences, 1992). A PSR mutató a következőképpen van definiálva:

Tegyük fel, hogy n iterációnk van m láncban, és θ_{ij} a paraméter értéke az i -edik iterációban, a j -edik láncban, ahol $i = 1, \dots, n, j = 1, \dots, m$. Az egyes láncokon belüli és láncok közötti szórás a következőképpen számolható:

$$W = (1/m) \sum_{j=1}^m (1/n) \sum_{i=1}^n (\pi_{ij} - \bar{\pi}_{\cdot j})^2$$

$$B = (1/(m-1)) \sum_{j=1}^m (\bar{\pi}_{\cdot j} - \bar{\pi}_{\cdot\cdot})^2,$$

ahol

$$\bar{\pi}_{\cdot j} = (1/n) \sum_{i=1}^n \pi_{ij},$$

$$\bar{\pi}_{\cdot\cdot} = (1/m) \sum_{j=1}^m \bar{\pi}_{\cdot j}.$$

A konvergenciát a PSR (lehetséges skála redukciós tényező) mutatóval határozzuk meg, melynek képlete:

$$PSR = \sqrt{\frac{(W + B)}{B}}$$

6. egyenlet: PSR

Gelman et al. (2004) javaslata szerint a PSR értéke legyen kisebb vagy egyenlő 1.1-nél minden paraméterre. A konvergencia akkor teljesül, ha a láncok közötti szórás kisebb, mint a láncokon belüli szórás (Muthén & Asparouhov, 2010).

A strukturális egyenletmodellek természetüknél fogva magukban foglalják a paraméterek meghatározását és becslését, az értékelést és az ellenőrzést, valamint a különböző modellek

közötti összehasonlítást is. A becsült paraméterek alapján a következő modellilleszkedést mérő mutató meghatározható.

Posteriori prediktív teszt (PPP)

Ebben az esetben egy posteriori prediktív p-értéket számítanak, amely összehasonlítja a megfigyelt valószínűségi arány tesztet a modell posterior eloszlása alapján létrehozott megfelelő tesztel. Legyen y^r azon elméleti értékek összessége, amelyeket a módosított kutatási modellből származtatunk: ez a becsült modellegyenletből kapott becsült y a megfigyelt változókra. Ekkor:

$$p(y^r | y) = \int p(y^r | \theta) p(y | \theta) p(\theta) d\theta.$$

7. egyenlet: PPP

A 7. egyenlet azt jelenti, hogy a jövőbeli megfigyelt értékek eloszlása az adott valós adatok mellett megegyezik a jövőbeli értékek eloszlásával, amelyet a modell ismeretlen paraméterei alapján, azok posterior eloszlásával súlyozva kapunk. A posteriori prediktív teszt azt mondja ki, hogy az elméleti adatoknak vagy értékeknek egyezniük kell a megfigyelt értékekkel, ha a modell illeszkedik a valós adatokhoz. Így egy p-érték jön létre az $LRT(y, \theta^s)$ valószínűségi arány statisztika alapján, ahol θ^s a paraméterek vektora, beleértve a megfigyelt adatok várható értékét, kovarianciamátrixát, és tartalmazza az $LRT(y^r, \theta^s)$ statisztikát is, amelyet a módosított modellből származtatott elméleti értékek alapján számítunk.

$$LRT(y, \theta^s) = 2[\log p(y | \hat{\theta}^{full}) - \log p(y | \theta^s)]$$

$$p - \text{érték} = P(LRT(y^r | \theta^s) > LRT(Y | \theta^s))$$

8. egyenlet: PPP p-érték

A körülbelül 0,5 körüli p-értékek azt jelzik, hogy a modell illeszkedik az adatokhoz, míg a nulla körüli p-értékek gyenge vagy nem létező illeszkedést jeleznek (Merkle & Rosseel, 2018).

A marginális log-likelihood (MargLogLik)

A bayesi statisztikában a marginális likelihood központi szerepet játszik az adatok illeszkedésének globális értékelésében, mutatja meg, hogy a teljes modell (prior + likelihood) mennyire képes globálisan megmagyarázni az adatokat. E mennyiséget gyakran logaritmikus skálán értékeljük:

$$\log p(y) = \log \int p(y | \theta) p(\theta) d\theta$$

9. egyenlet: MargLogLik

A MargLogLik szerepe elsősorban modellösszehasonlítás esetén fontos. Két vagy több modell közül az a preferált, amelyiknek a marginális valószínűsége nagyobb (log-skálán kevésbé negatív), feltéve, hogy ugyanazokkal az adatokkal dolgozunk.

1. Eredmények

Változószelekció

A termékenységi ráta alakulását potenciálisan befolyásoló számos társadalmi, gazdasági és politikai tényező figyelembevételével nagy változókészlet állt rendelkezésre (ld. Adatok fejezet). A bayesi strukturális egyenletmodell (BSEM) robusztus és értelmezhető felépítéséhez Lasso-regresszióval előzetes változószelekciót alkalmaztam, amely képes automatikusan nullára szorítani a kevésbé informatív változók együtthatóit, így kizárva azokat a modellből.

Először sztenderdizáltam a változókat (minden változónak 0 lesz az átlaga, és 1 a szórása), egyrészt kiküszöbölve a mértékegységeket és a nagyságrendi eltéréseket, másrészt a sztenderdizált modell együtthatói egymással összehasonlíthatók (ha X_j ceteris paribus a saját szórásával nő, Y várhatóan a saját szórásának β_j -szorosával nő).

Az L1-regularizációs paraméter optimalizálása 10-szeres keresztvalidációval történt, a minimális átlagos négyzetes hiba (MSE) alapján kiválasztott λ érték mellett, így a legtöbb információt megőrző, de szelektív modellt kaptam.

A Lasso szelekció eredményeként a következő magyarázóváltozók kerültek be a bayesi strukturális egyenletmodellbe:

- Közjavak és közszolgáltatások: Kormányzati költség rekreációra, kultúrára és vallásra;
Politikai stabilitás index

- Gazdasági ciklus: HICP; Fiatalok foglalkoztatási rátája; egy főre jutó GDP növekedés
- Nők helyzete: Nők munkaerőpiaci részvétele; Felsőoktatásban részesült nők aránya, Első házasságok aránya
- Ingatlanpiac: Lakásárindex; Jövedelemből az albérletre költött arány
- Életszínvonal: Nem várt kiadásokra való képtelenség; Szegénységből meneküléshez szükséges órák száma; Népsűrűség; Gyermekhalandóság; Várható élettartam
- Szabadidő: Minimum szabadság; Fizetett apaszabadság; Fizetett szülési szabadság
- Állami támogatás: Adópolitikai hatás; Kormányzati költség lakhatási támogatásra; Szociális védelmi kiadások; GMI

BSEM modell

A lasso módszerrel szelektált magyarázóváltozók termékenységre gyakorolt hatását bayesi strukturális egyenletmodellezéssel (Bayesian Structural Equation Modelling, BSEM) vizsgáltam meg a blavaan csomag¹³ segítségével. A modell 126 megfigyelést tartalmazott, és a becslést MCMC (Markov Chain Monte Carlo) módszerrel hajtottam végre, normál priorral (mean=0, sd=10) specifikált regressziós koefficiensek mellett, az iterációk száma 4000. A teljes modell 15 paramétert becsült.

Az országok közötti heterogenitás kontrollálása céljából dummy változókat hoztam létre az országokra vonatkozóan, s a referenciaország Ausztria lett. Több modellt is el kellett vetnem az abban tapasztalható súlyos multikollinearitás miatt, amelyet VIF-teszttel mértem le. Emiatt végül Magyarország és Szlovákia dummy változó nem került be a végső modellbe, mivel e két ország szerepeltetése redundanciát eredményezett. Az idő (év) változót sem illesztettem be a modellbe, mivel az egyértelműen nem bizonyult szignifikánsnak (~0 a becsült hatása), és az sem egyértelmű, hogy lineáris trend feltételezése indokolt lenne. Kezdeti próbálkozásként látens változók bevonásával is kísérleteztem (pl. gazdasági környezet, közszolgáltatások), azonban az alacsony elemszám (N=126) miatt a modell instabilitást mutatott, így ezen komplexebb specifikációkat nem tudtam megbízhatóan értelmezni. Ezért egy teljesen megfigyelt változókat szerepeltető (observed variables-only) BSEM modellt alkalmaztam (lsd. Melléklet).

Mivel feltételezhető, hogy az iskolázottság termékenységre gyakorolt hatása nem azonnali (Kravdal O. , 2002), a modellbe az iskolázottsági szint idősoros késleltetéseinek bevonását is megvizsgáltam. Kezdetben az iskolázottsági változó (felsőfokú végzettséggel rendelkező nők

¹³ R programozási nyelv egy csomaga

aránya) öt időkésleltetett értékét (lag1–lag5) szerepeltettem a modellben, hogy megragadhasam a késleltetett hatásokat. Azonban a változók között nagyon erős multikollinearitás jelentkezett, mivel az egymást követő évek oktatási adatai kis mértékben változnak, így erősen korrelálnak egymással. Ez a jelenség instabillá tette a becsült regressziós együtthatókat, csökkentve a modell megbízhatóságát. A probléma kezelésére az öt késleltetés egyszerű mozgóátlagát képeztem, így egyetlen, aggregált mutatót (Education_ma5) használtam a modellben. Ez az átalakítás eliminálta a multikollinearitást, így a modell stabillá vált és megőrizte a késleltetett hatások fő információtartalmát, miközben egyszerűsítette a becslést. Ezáltal a modell megbízhatóbban tükrözi az iskolázottság hosszabb távú hatását a termékenységre.

Ezek után több modellt is lefuttattam (ld. Melléklet 4. táblázat), majd a *MargLogLik* és a *PPP* mutatók alapján kiválasztottam a legjobb illeszkedésű modellt, amelyet lentebb részletesen ismertetek.

A láncok konvergenciáját az R-hat (PSR) értékek 1.000 körüli alakulása igazolja, amely minden változó esetében kiváló konvergenciát jelez: a láncok konvergáltak, megbízható a becslés. A modellem esetében a *MargLogLik* érték –143,386, amely a tesztelt modellek közül a legjobb illeszkedést mutatta.

A Bayes-i strukturális egyenletmodellek illeszkedésének értékelésére szolgáló Posterior Predictive p-érték (PPP) ideálisan 0.5 körül alakul: ilyenkor a modell által szimulált és a valós megfigyelések között nincs szisztematikus eltérés. Modellem PPP értéke 0.496, ami optimálisnak tekinthető, és megerősíti a modell szempontjából kielégítő globális illeszkedést. Az alternatív modellekhez viszonyítva ez rendelkezett a legjobb illeszkedési mutatókkal, ezért az értelmezést erre a modellre alapozom.

Az alábbi táblázatban (ld. 2. táblázat) láthatók a magyarázóváltozók¹⁴ becsült értékei (átlag), konfidenciaintervallumai és posterior szórásai. A konfidenciaintervallumok 95%-os megbízhatósági szinttel értendők.

¹⁴ A 2. táblázatban található változók nevei magyar fordításban megtalálhatóak a Mellékletben (3. táblázat)

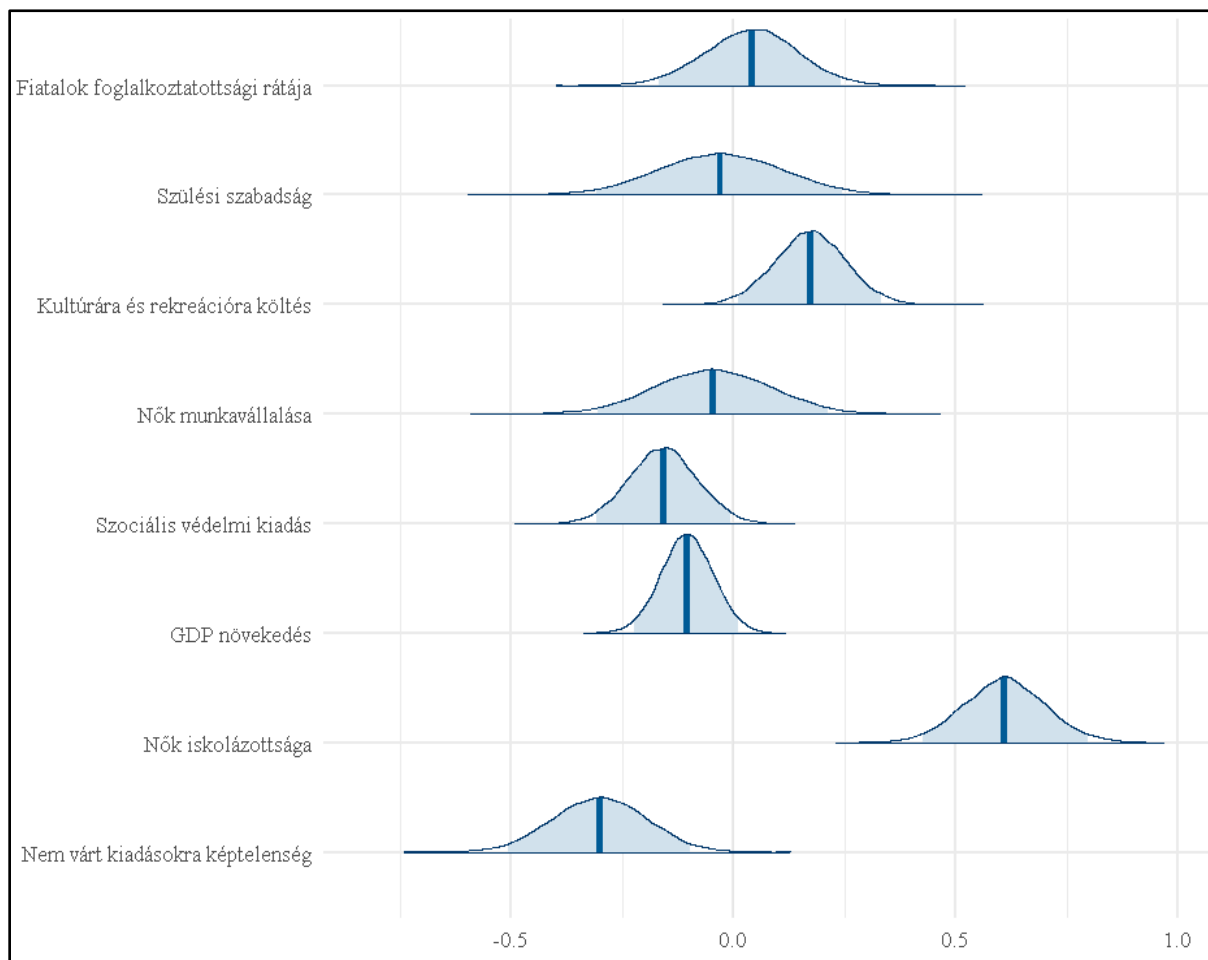
Változó	Becsült érték	Posterior szórás	Alsó intervallumhatár	Felső intervallumhatár	PSR
Paid_lev_mthrs	-0.030	0.141	-0.309	0.244	1.000
Family_support	-0.159	0.077	-0.309	-0.008	1.000
Inability	-0.301	0.106	-0.508	-0.097	1.000
Education_ma5	0.609	0.093	0.426	0.796	1.000
GDP_growth	-0.105	0.059	-0.222	0.010	1.000
Youth_employmnt	0.039	0.104	-0.168	0.241	1.000
Labor	-0.047	0.132	-0.305	0.211	1.000
Spendng_n_cltr	0.172	0.081	0.011	0.331	1.000
Estonia	0.473	0.281	-0.077	1.021	1.000
Czechia	0.891	0.196	0.507	1.266	1.000
Lithuania	0.223	0.391	-0.563	0.986	1.000
Latvia	0.697	0.348	0.007	1.371	1.000
Poland	-1.089	0.458	-1.991	-0.200	1.000
Slovenia	0.186	0.311	-0.438	0.787	1.000

2. táblázat: BSEM modell

Forrás: saját szerkesztés

Mivel minden változót standardizáltam (z-score transzformáció), az együtthatók közvetlenül összehasonlíthatóak, és a legnagyobb abszolút értékkel rendelkező regressziós koefficiensnek van a legnagyobb hatása a termékenységi rátára.

Ezen ábrán (lsd. 4. ábra) pedig a posterior eloszlások láthatók:



4. ábra: Posterior eloszlások

Forrás: saját szerkesztés

A magyarázóváltozók közül a legnagyobb együttható értéke a felsőoktatásban részesült nők arányának van: 0,61. A konfidenciaintervallum relatíve szűk, s egyértelműen szignifikáns a változó 95%-os megbízhatósági szinten. Ezen eredmény szerint ceteris paribus az elmúlt 5 év átlagos iskolázottsági szintjének növekedése magasabb termékenységi szintet okoz. Lefuttattam ezután egy within (fixed effects) regressziót is, aminek eredménye hasonló lett, minden szokásos szignifikancia szinten szignifikáns és pozitív az együtthatója. Ez kvázi egy robusztusságvizsgálat: a klasszikus frekventista panelökonometria megoldása arra a modellre, amit bayesi SEM-mel becsültem. Mindkét megközelítés ugyanoda vezet, így a hatás robusztus a modell választására.

Mindez ellentmond a korábbi bevett elméletnek, miszerint az iskolázottság növekedésével csökken a termékenység.

A korábbi elméletek, különösen a neoklasszikus családi közgazdaságtan (Becker, 1960) hangsúlyozták, hogy a nők iskolázottságának növekedése a gyermekvállalás alternatív költségének emelkedéséhez vezet, így negatív kapcsolat várható az iskolázottság és a termékenység között. Úgy vélem azonban, hogy Kelet-Európában van alapja annak, hogy ez az összefüggés nem feltétlenül érvényesül lineárisan vagy negatív irányban. Ennek háttérében több tényező állhat:

1. Erősebb hozzáférés az erőforrásokhoz és információhoz: Az iskolázottabb nők tudatosabban képesek kihasználni a gyermekvállalást támogató intézkedéseket, például a családtámogatási rendszert, a GYED/GYES lehetőségeit, vagy a bölcsődei infrastruktúrát (Rindfuss, Guilkey, Morgan, & Kravdal, 2010).
2. Késleltetett, de nem elmaradó gyermekvállalás: A magasabb iskolázottsággal bíró nők gyakran idősebb korban vállalnak gyermeket, miután már elérték szakmai céljaikat. Ez nem termékenységcsökkenést, hanem kitolódott gyermekvállalást eredményez (Bhrolcháin & Beaujouan, 2012).
3. Társadalmi értékváltás és intézményi alkalmazkodás: A nők fokozott jelenléte a munkaerőpiacon fokozza a keresletet a családbarát munkahelyek és gyermekintézmények iránt, ami közvetetten támogatja a termékenységet. A társadalmi normák is változnak: a gyermekvállalás és a karrier nem feltétlenül kizáró fogalmak (Sobotka, 2004).

Összességében úgy gondolom, hogy az iskolázottabb nők jobban tudnak optimalizálni az elérhető támogatási lehetőségek és családbarát munkahelyek (pl. részműködés foglalkoztatási lehetőségek) között.

Abszolút értékben a 2. legnagyobb együtthatója a nem várt kiadásokra való képtelenség társadalmi arányának volt. Ez szintén statisztikailag szignifikáns 95%-on és negatív. A súlyos anyagi nehézségekkel küzdők arányának növekedése csökkenti a termékenységet, tehát a mélyszegénységben élők kevésbé vállalnak gyermeket a kelet-európai régióban. Ez az eredmény összhangban áll a gazdasági bizonytalanságról szóló elméleti és empirikus kutatásokkal, amelyek szerint az anyagi kiszolgáltatottság a gyermekvállalás elhalasztásához vagy elmaradásához vezethet. Az alábbi okok magyarázhatják ezt a negatív kapcsolatot:

1. Egzisztenciális bizonytalanság: Az anyagi nehézségek közvetlenül csökkentik a családalapításra való hajlandóságot, különösen akkor, ha a jövőbeli jövedelmi kilátások is kedvezőtlenek. A gyermekvállalás költségei jelentős terhet róhatnak az amúgy is sérülékeny háztartásokra (Kreyenfeld, 2010).
2. Rövid távú túlélési preferenciák: A tartós anyagi nélkülözésben élők gyakran a mindennapi túlélésre koncentrálnak, így a gyermekvállalás, amely hosszú távú tervezést igényel, háttérbe szorul. Az időzítés kérdése helyett maga a gyermekvállalás kerül megkérdőjelezésre (Comolli, 2017). Az anyagi szorongás és tartós stressz negatívan hat a párkapcsolati stabilitásra és a családtervezési döntésekre is. Ez különösen igaz Kelet-Európára, ahol a gazdasági bizonytalanság a rendszerváltás után évtizedekig meghatározó tényező volt.

A kulturális és szabadidős kiadások emelkedése pozitív és statisztikailag szignifikáns kapcsolatban áll a termékenységgel ($\beta = 0.172$). Ez arra mutat rá, hogy azon országokban, ahol nagyobb hangsúlyt kap a közösségi élet és a szabadidős infrastruktúra fejlesztése, kedvezőbb társadalmi környezet alakul ki a gyermekvállaláshoz. Az ilyen kiadások fokozhatják az életminőséget, a társadalmi kohéziót, ami ösztönözheti a családalapítást.

Ezzel szemben a családi szociális védelemre fordított kiadások GDP-arányos szintje negatív és szignifikáns kapcsolatot mutat a termékenységgel ($\beta = -0.159$), bár nem annyira erős a hatás, illetve a megbízhatóság sem olyan, mint az előző változók esetén. Az eredmény kontraintuitív, azt várnánk a szakirodalom alapján is, hogy a támogatások növekedése növeli a termékenységet. Ez arra utal, hogy a családtámogatások nem mindig érik el a célzott hatást, különösen, ha azok inkább kompenzáló, mint ösztönző jellegűek – vagyis a magas támogatások inkább következményei a demográfiai problémáknak, mintsem megoldásai. A másik probléma feltételezhetően magával a mutatóval van, ami minden családi szociális védelmi kiadást tartalmaz (pl. gyermekek kötelező oltakozása, ingyenes tankönyv) és ebbe sokszor olyan tételek is beletartoznak, amik nem célozzák a gyermekvállalás elősegítését. A nagyobb általános szociális védelmi kiadásoknak (pl. általános munkanélküli segély) olyan hatása is lehet, hogy arra ösztönözhetik a nőket, hogy ne szüljenek, mivel enélkül is számos „bőkezű” támogatási forma elérhető számukra.

A nők munkavállalása, a GDP-növekedés, a szülési szabadság, valamint a fiatalok foglalkoztatási rátája se bizonyult szignifikánsnak érdekes módon.

Az országfixhatásokra kontrollálás eredményeként Csehország és Lettország mutatott pozitív és statisztikailag szignifikáns eltérést a referenciához képest, míg Lengyelország esetében negatív, szignifikáns hatás érvényesült; Szlovénia és Litvánia nem szignifikáns a referenciaországhoz (Ausztria) képest.

Limitáció

Modellem limitációja, hogy relatíve kicsi a minta: $N=126$. Több ország esetén nagy az aránya a hiányzó adatoknak, így nem tudtam lefedni a teljes kelet-európai régiót, csupán annak egy részét. Emellett a családtámogatási rendszerekről nem érhető el egységes, átfogó adatbázis, így több változó esetén nincs megfelelő adat (pl. családi pótlék összege, egyéb normatív támogatások...stb.). Ha ezek az adatok visszamenőlegesen rendelkezésre is állnak azok is csak a nemzeti statisztikai hivatalok adatbázisaiban, amit óriási munka kibányászni, erre pedig nem volt elég időm és erőforrásom jelen dolgozat keretein belül. Értékes munka lenne a későbbiekben egy ilyen adatbázist összeállítani. Sajnos a kezdeti kutatási kérdéseimre így nem tudtam teljes körű választ adni, főleg az állami beavatkozások kérdéskörét nem tudtam kellően körüljárni.

VII. Konklúzió

Kutatásomban a különböző 2006-2019 közötti állami beavatkozások (pl. szülési szabadság, családi szociális kiadások), valamint a különböző társadalmi és gazdasági tényezők hatását vizsgáltam a termékenységi rátára bayesi strukturális egyenletmodell (BSEM) segítségével Kelet-Európa 9 EU-s országában. Eredményeim alapján a nők felsőoktatásban való részvétele befolyásolja leginkább a gyermekvállalást és azt találtam ellentétben a szakirodalommal, hogy pozitív az összefüggés a két változó között, tehát minél magasabb egy országban a nők iskolázottsága annál magasabb a termékenység. Úgy vélem e folyamat mögött az áll, hogy az iskolázottabb nők jobban tudnak optimalizálni az elérhető támogatási lehetőségek és munkahelyek között. Megerősítést nyert továbbá, hogy a mélyszegénységben élők kevésbé hajlamosak gyermeket vállalni. Ezen eredmények alapján tehát a kormányzatoknak fontos lenne ösztönözni a továbbtanulást, hogy a jobb anyagi körülmények között élő értelmiség bővülésével a gyermekvállalási kedv növekedjen, valamint a mélyszegénységben élők számára

szükséges lenne célzott szociális programokat indítani, hogy az egzisztenciális bizonytalanság ne állhasson a gyermekvállalás útjába.

Kutatásomnak azonban több limitációja is akad, főképp a kis minta és több fontos változó hiánya (a hiányzó értékek magas aránya miatt) torzíthatja az eredményeket. Úgy vélem, hogy kiemelkedő fontossággal bírna egy egységes, átfogó multinacionális adatbázis szerkesztése a különböző családpolitikai rendszerekről.

A megfelelően kialakított és hatékonyan működő kormányzati családpolitikák hosszú távon jelentős hatást gyakorolnak a társadalmi jólétre. Ennek következtében kiemelten fontos, hogy a kormányzat és a szakpolitikai döntéshozók folyamatosan értékeljék és finomítsák az otthonteremtési politikákat annak érdekében, hogy azok megfeleljenek a lakosság és a gazdaság változó igényeinek és kihívásainak.

Melléklet

Irodalomjegyzék

- Andersson, G., Duvander, A.-Z., & Hank, K. (2004). Do child care characteristics influence continued childbearing in Sweden? An investigation of the quantity, quality and price dimension. *Journal of European Social Policy*, 14, 407–418.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.
- Becker, G. (1960). An Economic Analysis of Fertility. *National Bureau of Economic Research*, 209-240.
- Becker, G. (1981). *A Treatise on the family*. Cambridge, MA: Harvard University Press.
- Berghammer, C., & Beaujouan, E. (2019). The Gap Between Lifetime Fertility Intentions and Completed Fertility in Europe and the United States: A Cohort Approach. *Population Research and Policy Review*, Volume 38, pages 507–535.
- Bhrolcháin, M., & Beaujouan, E. (2012). Fertility postponement is largely due to rising educational enrolment. *Population Studies*, 66(3):311-27.
- Björklund, A. (2006). Does family policy affect fertility? *Journal of Population Economics*, Volume 19, pages 3–24.
- Black, D., Kolesnikova, N., Sanders, S., & Taylor, L. (2013). Are Children “Normal”? *Rev Econ Stat.*, 95(1):21–33.
- Blanchet, D., & Ekert-Jaffé, O. (1994). The Demographic impact of family benefits: evidence from a micro-model and from macro-data. In E. J.-O. N, *The Family, the Market and the State in Ageing Societies* (old.: 79-104). Oxford: Clarendon Press.
- Breton, D., & Prioux, F. (2005). Deux ou trois enfants? Influence de la politique familiale et de quelques. *Population*, 4(Juillet-aout), Vol. 60, p. 415-445.
- Caudill, S., & Mixon Jr., F. (1993). A note on the effects of AFDC payments on birthrates. *International Review of Economics and Business*, 40(4): 379-384.
- Chesnais, J.-C. (1996). Fertility, family, and social policy. *Population and Development Review*, 22(4), 729–739.
- Cigno, A. (1991). *Economics of the family*. Oxford, England: Clarendon Press.
- Comolli, C. (2017). The fertility response to the Great Recession in Europe and the United States: Structural economic conditions and perceived economic uncertainty. *Demographic Research*, Volume 36, Article 51, Pages 1549-1600.
- Del Boca, D., Aaberge, R., Colombino, U., Ermisch, J., Francesconi, M., Pasqua, S., & Strom, S. (2003). *Labour market participation of women and fertility: The effect of social policies*. Report for the V European Conference.

- Doepke, M. (2015). GARY BECKER ON THE QUANTITY AND QUALITY OF CHILDREN. *Journal of Demographic Economics* , Volume 81 , Issue 1.
- Ekert, O. (1986). Effets et limites des aides financières aux familles: Une expérience et un modèle. *Population*, 2, 327–348.
- Ekert-Jaffé, O., Joshi, H., Lynch, K., Mougin, R., & Rendall, M. (2002). Fertility, timing of births and socioeconomic status in France and Britain: social policies and occupational polarisation. *Population (French Edition)*, 57:475-507.
- Ermisch, J. (1988). The econometric analysis of birth rate dynamics in Britain. *The Journal of Human Resources*, 23(4), 563–576.
- Gauthier, A. H. (2007). The impact of family policies on fertility in industrialized countries: a review of the literature. *Population Research and Policy Review*, Volume 26, pages 323–346.
- Gauthier, A., & Hatzius, J. (1997). Family Benefits and Fertility: An Econometric Analysis. *Population Studies*, 51:3, 295-306.
- Gelman, A., & Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7 (4):457–511.
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian Data Analysis. 2nd ed.* London: Chapman & Hall, Ltd.
- Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 (6):721–41.
- Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Goldthorpe, J. (2000). *On sociology; Numbers, narratives, and the integration of research and theory*. Oxford, England: Oxford University Press.
- Hank, K., & Kreyenfeld, M. (2003). A multilevel analysis of childcare and women's fertility decisions in Western Germany. *Journal of Marriage and the Family*, 65(3), 584–596.
- Hoem, J. (1993). Public policy as the fuel of fertility: Effects of a policy reform on the pace of childbearing in Sweden in the 1980s. *Acta Sociologica*, 36(1), 19–31.
- Hoem, J., Prskawetz, A., & Neyer, G. (2001). Autonomy or conservative adjustment? The effect of public policies and educational attainment on third births in Austria, 1975–96. *Population Studies*, 55(3), 249–261.
- Hyatt, D., & Milne, W. (1991). Can public policy affect fertility? *Canadian Public Policy/Analyse de Politiques*, 27(1), 77–85.
- Jackman, S. (2000). Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo. *American Journal of Political Science*, 44 (2):375–404.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Hove and London: Scientific Software International.

- Kaplan, D., & Depaoli, S. (2012). "Bayesian Structural Equation Modeling." In: *Handbook of Structural Equation Modeling*, edited by R. Hoyle, 650–673. New York: The Guilford Press.
- Kontakos, S., & Tsairidis, C. (2024). Bayesian structural equation modeling. A detailed Review and an application in the intention to "Brain Drain" of the graduates of the Social Administration Department of Democritus University of Thrace. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 11(1), 1–30.
- Kovács, L. (2021). *Változószelektációs algoritmusok vizsgálata általánosított additív modellekben*. Budapest: Budapesti Corvinus Egyetem .
- Kravdal, Ø. (1996). How the local supply of day-care centers influences fertility in Norway: A parity-specific approach. *Population Research and Policy Review*, 15(3), 201–218.
- Kravdal, O. (2002). Education and fertility in Sub-Saharan Africa: individual and community effects. *Demography*, Volume 39, pages 233–250.
- Kreyenfeld, M. (2010). Uncertainties in Female Employment Careers and the Postponement of Parenthood in Germany . *European Sociological Review*, Volume 26, Issue 3, Pages 351–366.
- Lee, R. (2015). Becker and the Demographic Transition. *Journal of Demographics Economics*, 1;81(1):67–74.
- Lee, S.-Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. Chichester: John Wiley & Sons.
- Lee, S.-Y., & Song, X.-Y. (2004). Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57 (Pt 1):131–50.
- Link, W., & Eaton, M. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3 (1):112–15.
- Magnus, G. (2008). *The Age of Aging: How Demographics are Changing the Global Economy and Our World* . John Wiley & Sons Ltd.
- Merkle, E., & Rosseel, Y. (2018). blavaan: Bayesian Structural Equation Models via Parameter Expansion . *Journal of Statistical Software*, 85 (4):1–30.
- Muthén, B., & Asparouhov, T. (2010). Bayesian SEM: A More Flexible Representation of Substantive Theory. *Psychological Methods*, 17 (3):313–35.
- Pacalova , H., Engelhardt, H., Morgan, S., & Diprete, T. (2003). Do Cross-National Differences in the Costs of Children Generate Cross-National Differences in Fertility Rates? *Population Research and Policy Review*, 22(5–6), 439–477.
- Press, S. (2003). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. New Jersey: John Wiley & Sons.
- Rindfuss, R., Guilkey, D., Morgan, S., & Kravdal, O. (2010). Child-Care Availability and Fertility in Norway. *Population and Development Review*, 36(4):725–748.

- Ronsen, M. (2004). Fertility and public policies—Evidence from Norway and Finland. *Demographic Research*, 10, 143–170.
- Sobotka, T. (2004). *Postponement of childbearing and low fertility in Europe*. Amsterdam: Dutch University Press.
- Szabó-Morvai, Á., Balás, G., Bördös, K., & Herczeg, B. (2019). *Evaluation of family policy measures and their impact on fertility*. Budapest: Hétfa Research Institute.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, Vol. 58, No. 1, pp. 267-288.
- Zhang, J., Quan, J., & Van Meerbergen, P. (1994). The Effect of Tax-Transfer Policies on Fertility in Canada, 1921-88. *The Journal of Human Resources*, Vol. 29, No.1, pp. 181-201.
- Zhang, T.-T., Cai, X.-Y., Shi, X.-H., Zhu, W., & Shan, S.-N. (2023). The Effect of Family Fertility Support Policies on Fertility, Their Contribution, and Policy Pathways to Fertility Improvement in OECD Countries. *International Journal of Environmental Research and Public Health*, 20(6):4790.
- Zhao, P., & Yu, B. (2006). On Model Selection Consistency of Lasso. *Journal of Machine Learning Research* 7, 2541-2563 .

<https://data.oecd.org/pop/fertility-rates.htm>

<https://ksh.hu/s/kiadvanyok/fenntarthato-fejlodes-indikatorai-2022/1-2>

<https://europa.eu/eurobarometer/surveys/detail/3112>

<https://ec.europa.eu/eurostat>

<https://www.worldbank.org/en/home>

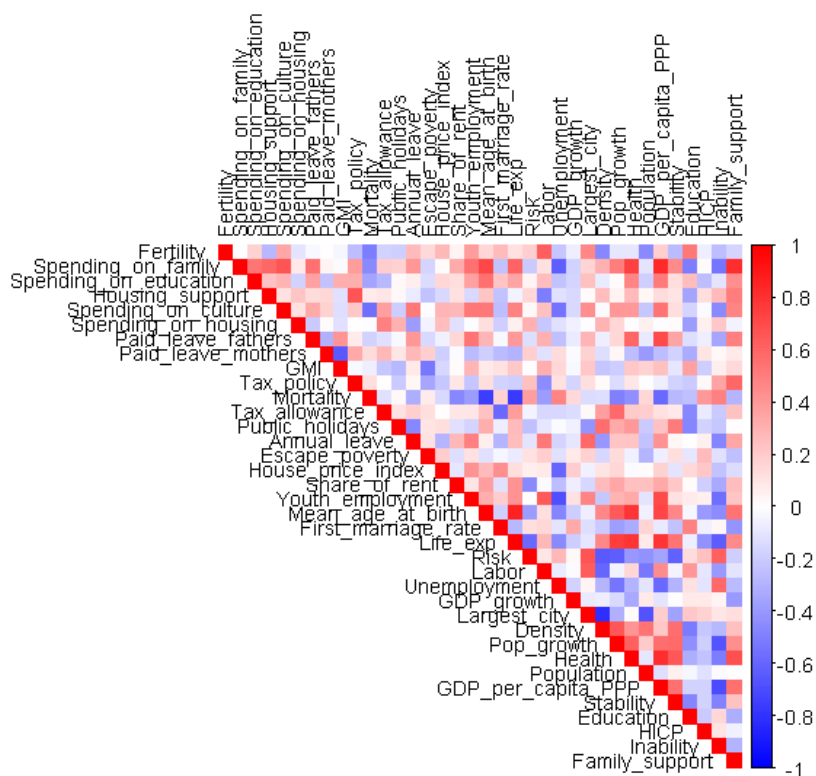
<https://www.oecd.org/en/data/datasets/social-expenditure-database-socx.html>

<https://demografia.hu/hu/tudastar/fogalomtar/38-teljes-termekenysegi-aranysszam>

<https://ksh.hu/s/kiadvanyok/fenntarthato-fejlodes-indikatorai-2022/1-2>

<https://www.ibm.com/think/topics/lasso-regression>

<https://www.ibm.com/think/topics/lasso-regression>



5. ábra: Korrelációs mátrix

Forrás: saját szerkesztés

Változó neve	Változó neve a modellben
Fizetett szülési szabadság	Paid_lev_mthrs
Szociális védelmi kiadások (GDP %-ában)	Family_support
Nem várt kiadásokra való képtelenség (társadalom %-a)	Inability
Felsőoktatásban részesült nők aránya	Education
Egy főre jutó GDP növekedés (évi %)	GDP_growth
Fiatalok foglalkoztatási rátája (%)	Youth_employmnt
Nők munkaerőpiaci részvétele (15+ életkorú nők %-ában)	Labor
Kormányzati költség rekreációra, kultúrára és vallásra (USD 2015 PPP/fő)	Spendng_n_cltr
Észtország	Estonia
Csehország	Czechia
Litvánia	Lithuania
Lettország	Latvia
Lengyelország	Poland
Szlovénia	Slovenia

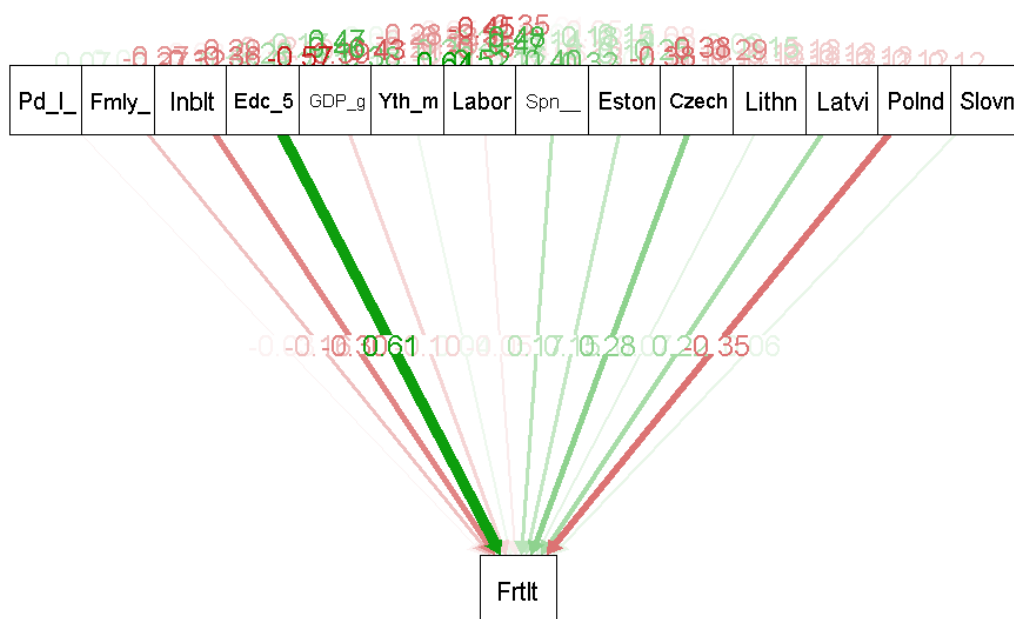
3. táblázat

Forrás: saját szerkesztés

	1. modell	2. modell	3. modell	4. modell	5. modell
Változó	Becsült érték				
Paid_lev_mthrs	0.173	-0.533	0.016	0.069	-0.030
Family_support	-0.281	-0.006	-0.167	-0.158	-0.159
Inability	-0.114	-0.237	-0.352	-0.359	-0.301
Education	0.471	0.563	0.462	0.620	0.609
GDP_growth	-0.158	-0.097	-0.121	-0.116	-0.105
Youth_employmnt	0.073	0.216	0.109	0.124	0.039
Labor	0.251	-0.163	-0.190	-0.163	-0.047
Spendng_n_cltr	0.348	0.354	0.206	0.234	0.172
Estonia	-	2.228	0.867	0.557	0.473
Czechia	-	1.968	0.789	0.849	0.891
Lithuania	-	1.223	0.721	0.498	0.223
Latvia	-	1.827	1.110	0.966	0.697
Poland	-	-0.529	-0.846	-0.879	-1.089
Slovenia	-	0.684	0.449	0.333	0.186
Hungary	-	1.341	-	-	-
Slovak Republic	-	2.083	-	-	-
Time	-	-	0.031	-	-
MargLogLik	-157.318	-154.598	-159.51	-154.305	-143.363
PPP	0.498	0.499	0.485	0.496	0.496

4. táblázat

Forrás: saját szerkesztés



6. ábra: LISREL ábra

Forrás: saját szerkesztés

Nyilatkozat

A dolgozat elkészítése során ChatGPT-t használtam az R-ben való programkódok elkészítéséhez (pl. ábrákhoz, modellekhez), ötletelésre, szakirodalom fordításra és forrásgyűjtésre.