

EGÉSZSÉGPOLITIKAI INTÉZKEDÉSEK HATÁSA EGÉSZSÉGÜGYI INDIKÁTOROKRA MAGYARORSZÁGON

Módszertani útmutató egyéni szintű adatokon alapuló becslésekhez
készítette: Elek Péter¹

1. Bevezetés

Az alábbi módszertani útmutatóban azt foglaljuk össze, hogy milyen lehetőségeket rejtenek az egyéni szintű adminisztratív adatok az egészségpolitikai intézkedések utólagos hatásvizsgálatában, és milyen szempontok szerint, milyen lépéseket követve kell elvégezni ilyen hatásvizsgálatokat. Az útmutató közvetlenül a 2006 és 2008 közötti egészségpolitikai átalakítások monitorozásához kapcsolódik, de célunk, hogy más hatásvizsgálatok elvégzéséhez is segítségül szolgáljon.

Az utólagos hatásvizsgálatok célja, hogy a beavatkozások tényleges hatásait elkülöníthessük az egyidőben zajló egyéb tényezők hatásaitól. Ezt egyéni szintű adatok alapján nagyobb pontossággal meg lehet tenni, mint aggregált adatok alapján, ráadásul az egyéni szintű adatok olyan összefüggésekre is rámutathatnak (pl. bizonyos csoportok nagyobb érintettsége), amelyekre az aggregált adatok nem képesek.

A hatásvizsgálatok módszertani kérdéseinek jó áttekintését adja Khandker és szerzőtársai (2010) műve, kifejezetten az egészségügy területén pedig Grun (2006) tanulmánya. Az ökonometriai technikák mélyebb megértéséhez Wooldridge (2002 és 2009) tankönyvei (magyar nyelven Ramanathan 2003) nyújthatnak fogódzót.

2. Hatásvizsgálat elvégezhetőségének feltételei

A hatásvizsgálat előfeltétele, hogy egyértelműen meghatározható legyen: mikor történt az intézkedés, kiket érintett, kiket nem, milyen időszakban lehetett (elméleti alapon) valamilyen hatása, és milyen kimenetekre. Ezen belül szükséges, hogy az intézkedés által érintettek köre időben vagy valamilyen megfigyelt ismérv alapján behatárolható legyen, a várt hatások mérhetőek legyenek. Végül, az intézkedésre, az érintettekre, a nem kezelt csoportra és a hatásokra vonatkozó adatok kellő mintanagysággal rendelkezésre kell álljanak.

3. Adatok

Adatforrások. A hatásvizsgálathoz szükséges adatok fő forrásai az esetszintű egészségügyi adatbázisok, amelyek minden esetre tartalmazzák:

- az ellátással kapcsolatos változókat (ápolási napok, diagnózisok, beavatkozások,

¹ Eötvös Loránd Tudományegyetem Közgazdaságtudományi Tanszék. A szerző köszönettel tartozik Scharle Ágotának az útmutató elkészítéséhez nyújtott hasznos szakmai tanácsaiért.

gyógyszerfelhasználás stb.),

- az ellátott személy egyedi, de személyazonosításra alkalmatlan azonosítóját,
- az ellátás helyét (település, intézmény, osztály) és ehhez kapcsolhatóan az intézmény, osztály finanszírozási és egyéb jellemzőit (pl. teljesítményvolumen-korlát, ágyszám),
- az egyén fontosabb demográfiai jellemzőit (életkor, lakóhely stb.) és ehhez kapcsolhatóan a lakóhelyet jellemző társadalmi-gazdasági változókat (pl. településtípus, kistérségi munkanélküliségi ráta stb.). Esetleg – bár ez technikailag nehezebben megvalósítható – az egyén más társadalombiztosítási adatbázisokból származó jellemzői is hozzákapcsolhatók az adatbázishoz (pl. pénzügyi transzferek igénybe vételére, munkaviszonyra vonatkozó adatok – ezek nyilvánvalóan összefüggnek az egészségi állapottal).

Adatok struktúrája. A fenti adatforrásokból származó mikroszintű adatok legtöbbször egyesített keresztmetszeti ("pooled cross section") típusúak, azaz különböző időszakokra és különböző emberekre vonatkoznak. (Természetesen lehetséges, de nem szükségszerű, hogy egy embert éppen több időszokban is megfigyelünk.) Ha azonos embereket figyelünk meg különböző időszakokban, akkor panel adatbázisról, ha pedig csak egy időpontból származik a mintánk, akkor egyszerű keresztmetszeti mintáról beszélünk.

4. A hatásvizsgálat lépései

4.1. Az intézkedés és a várt hatás behatárolása, változók típusai

Intézkedés. A hatásvizsgálat első lépéseként meg kell pontosan határozni a "kezelést" (intézkedést) és a várt hatást. Ha pontosan definiálható, hogy bizonyos egyének bizonyos időszakban megkapták a "kezelést", mások/máskor pedig nem, akkor az intézkedés 0-1 értéket felvevő (dummy) változóval jellemezhető. (Például a vizitdíj bevezetése esetén az érintett emberekre az érintett időszakban a változó 1-et, egyébként pedig 0-t vesz fel.) Máskor az intézkedés közvetlenül egy folytonos változót érint, ekkor ezt a változót érdemes szerepeltetni. (Például az ágyszám-csökkentési intézkedés esetén nem dummyt érdemes használni /csökkent-e az ágyszám vagy nem/, hanem az adott intézmény ágyszámának százalékos változását az elmúlt évben /vagy fél évben/, mert ez utóbbival valószínűleg pontosabban lehet mérni az indikátorra kifejtett hatást.) Hasonlóan a többi, az adatbázisban szereplő változó is lehet folytonos vagy diszkrét.

Várt hatás. Az intézkedés várt hatását valamilyen indikátor (eredményváltozó) javulásában vagy romlásában mérjük. Az eredményváltozó is lehet folytonos vagy diszkrét (ez utóbbi esetben általában 0-1 értéket felvevő dummy). A folytonos eredményváltozó speciális esete az időtartam-változó, amikor egy esemény bekövetkezéséig (halál, kórházba való visszavétel stb.) eltelt időt mérjük. Például kíváncsiak lehetünk arra, hogy a kórházból elbocsátott pszichiátriai betegek mennyi idő múlva kerülnek újra kórházba. Ekkor az eredményváltozó időtartam típusú, de készíthetünk belőle dummyt például úgy, hogy a 30 napon belül visszavett betegek arányát vizsgáljuk, amivel az orvosi szempontból releváns információt hatékonyan tudjuk egy 0-1 változóba sűríteni. (Az eredményváltozó egyéni szinten 1-et vesz fel, ha a beteget 30 napon belül visszavették, egyébként pedig 0 az értéke.)

Előfordul, hogy az eredményváltozó (vagy akár valamelyik magyarázó változó) cenzorált, azaz bizonyos mintabeli elemekre csak azt tudjuk róla, hogy egy adott küszöb felett (vagy esetleg alatt) helyezkedik el. Ez különösen gyakori időtartam-változóknál. Például azon elbocsátott betegek esetén, akik nem kerültek még vissza a kórházba, a visszavételig eltelt időről csak azt tudjuk, hogy egy bizonyos (egyéntől függő) küszöb felett van. A cenzorálásra az

elemzési technika alkalmazásakor majd figyelemmel kell lenni.

4.2 Ismerkedés az eredmény- és magyarázó változókkal, leíró statisztika, adattisztítás

Egyváltozós leíró statisztikai vizsgálat. Az ökonometriai elemzés előtt célszerű a változók leíró statisztikai vizsgálata (átlag, szórás, kvartilisek, legnagyobb és legkisebb érték, kiugró értékek, a változó hisztogramjának ábrázolása). A kiugró értékek elemzésével az esetleges elütések is kiszűrhetők és korrigálhatók. Ezenkívül további adattisztítás is szükséges lehet az adatok belső logikai konzisztenciájának biztosításához (ellentmondások kiszűréséhez). Külön meg kell vizsgálni a hiányzó értékeket (melyik változóból mennyi érték hiányzik és miért), mert később a regressziók becslése a legszűkebb adathalmazon fog történni (azaz ahol semelyik használt változó sem hiányzik). Sokszor a hiányzó adatok valójában nullákat takarnak, ekkor azokat ki kell egészíteni. Amennyiben azért hiányzik egy adat, mert a változó egy másik változó logaritmálásával állt elő, és az eredeti változóban nullák voltak (azaz a transzformált változó "értéke" $-\infty$), akkor érdemes lehet egy nagyon nagy abszolút értékű negatív számot beírni a helyükre, és úgy futtatni a regressziókat.

Adatok összefüggése. A leíró statisztikai elemzés következő lépéseként vizsgálnunk kell az eredményváltozó és az intézkedési változó nyers (más változók kiszűrése nélküli) összefüggését. Amennyiben az intézkedési változó dummy (0-1 értékű), akkor egyszerűen ábrázolhatjuk az eredményváltozó részátlagát az intézkedési változó 0 és 1 értéke esetén, vagy ugyanezeket a részátlagokat még egyéb változók (év, korcsoport stb.) szerinti bontásban is. Például a 2007 februárjától 2008 áprilisáig érvényben levő vízitdíjat nem kellett fizetniük a 18 év alattiaknak, ezért a vízitdíj intézkedési változó csak a felnőttekre és csak ebben az időszakban 1, egyébként pedig 0. Így első leíró statisztikai elemzésként érdemes az eredményváltozó felnőttekre és gyerekekre vonatkozó átlagát havonta ábrázolni. Az így készített ábrából rögtön látszik, ha a vízitdíj hatására az eredményváltozó felnőttekre és gyerekekre vonatkozó átlagértékei közötti különbség 2007-ben a többi évhez képest jelentősen megváltozott, ami (előzetesen) a vízitdíj hatását mutatná.

Folytonos intézkedési változó esetén például a változó magas, közepes és alacsony értékeire érdemes az eredményváltozó részátlagait kiszámítani és ábrázolni.

4.3 A hatás identifikálása

A változók definiálása után az adatok előzetes vizsgálatával kiderül, hogy az intézkedések hatása mi alapján azonosítható. Az intézkedések egy része időben szór csak (pl. nyilvános várólisták bevezetése), máskor azonban vannak kevésbé és jobban érintett csoportok, amivel a szórás a keresztmetszeti dimenzióban növekszik. Példa erre a már említett vízitdíj, vagy ha az ágyszámok csökkentése különböző nagyságú volt megyénként, akkor ez az időbeli dimenzióhoz képest többlet varianciát visz az intézkedési változóba.

Regressziós keret a hatás identifikálására. Ideális (de az egészségpolitikai hatásvizsgálatokban szinte soha nem előforduló) esetben az intézkedéssel érintett és nem érintett megfigyelések az intézkedés kivételével minden egyéb tulajdonságukban megegyeznek. (Ez a helyzet a természettudományokban, tervezett kísérletekkor.) Ekkor az intézkedés hatása egyszerűen becsülhető az érintett és nem érintett megfigyelések eredményének különbségével. A gyakorlatban azonban az intézkedés így kiszámolt látszólagos hatásából ki kell szűrni azt a különbséget, ami a két csoport tulajdonságainak szisztematikusan eltéréseiből származik. Figyelembe kell tehát venni, hogy az intézkedés mellett általában más

tényezők is hatnak eltérően az érintett és nem érintett csoportokra. Regressziós keretben, az intézkedési változó mellett egyéb magyarázó változókat szerepeltetve kiszűrhetjük ezen „zavaró” tényezők hatását.

Fontos tehát, hogy egy regresszióban a becsült paramétereket (speciálisan például az intézkedési változó paraméterét) csak akkor lehet tényleges *ok-okozati összefüggésként (hatásként)* értelmezni, ha a szereplő magyarázó változók *exogének*, azaz minden egyéb releváns tényezőt kiszűrtünk a modellben. A többi tényező kiszűrése nem mindig egyszerű feladat, több nem nyilvánvaló okból is csorbulhat az exogenitás, azaz léphet fel endogenitás. Ezekből néhányat (nem megfigyelhető változók, szelekciós torzítás) az 5. fejezet tárgyal.

Megjegyezzük, hogy léteznek egyéb szűrő eljárások is, például a diszkontinuitás-modellek és a párosítási módszerek. Ezekkel itt nem foglalkozunk részletesen, de rövid leírásuk megtalálható a Függelékben (vagy részletesebben lásd pl. Khandker és szerzőtársai 2010). Az endogenitás problémája egyébként paneladatok használatával vagy keresztmetszeti adatok esetén az ún. instrumentális változók módszerével jobban kezelhető, mint egyszerű keresztmetszeti regresszióban. (A paneladatok azért előnyösebbek, mert explicit módon tudunk kontrollálni az egyének nem megfigyelhető tulajdonságaira.) Ezeknek a modern ökonometriai módszereknek a rövid leírása is megtalálható a Függelékben (illetve lásd pl. Wooldridge 2002 és 2010 tankönyveit).

Átviteli mechanizmus. Sok esetben az intézkedések csak közvetetten hatnak az eredményváltozóra, ezért szükség lehet a közvetlenül ható tényezők (az intézkedés és az eredményváltozó közötti „kapocs”) feltárására. Az átviteli mechanizmus azokat a folyamatokat írja le, amelyekeken keresztül az intézkedés hat az indikátor alakulására (például az ellátási területek átrendezése a bejártott orvos-beteg kapcsolatok felbomlásán keresztül hathat.) Ezeket a mechanizmusokat orvos-szakmai találkozások során érdemes feltérképezni, meghatározni. A mechanizmusok számszerűsítésére előfordulhat, hogy úgynevezett proxy-t kell találni, azaz egy olyan – folytonos vagy diszkrét – változót, amelyről azt gondoljuk, hogy az átviteli mechanizmust jól modellezi (például az orvos-beteg kapcsolatok változékonyságát az intézmény-váltásokkal modellezhetjük). A jó proxy tehát egy olyan számszerűsíthető változó, amely szakmai ismereteink, tapasztalataink szerint erősen korrelál a modellezendő közvetlen (de nem megfigyelt) változóval.²

Ha találunk tehát egy olyan változót, amely megfelelő közbülső „kapocsként” szolgál az intézkedési változó és a végső indikátor között, akkor az intézkedés hatását kiszámolhatjuk két együtthető (az intézkedés közbülső változóra illetve a közbülső változó végső indikátorra kifejtett hatása) szorzataként. Ez a közvetett eljárás azonban csak két feltétel együttes teljesülésekor alkalmazható. Egyrészt, az intézkedés csak a kiválasztott közbülső változón (változókon) keresztül hathat a végső indikátorra (különben alulbecsülnénk az intézkedés hatását). Másrészt a kiválasztott közbülső változónak exogénnek kell lennie a végső eredményváltozó egyenletében, azaz minden egyéb tényezőt ki kell tudnunk szűrni (precízebben fogalmazva, a modell hibatagjának korrelálatlannak kell lennie a magyarázó változókkal). Ha ez utóbbi feltétel nem teljesül, akkor a közbülső változó végső indikátorra kifejtett hatását nem tudjuk számszerűsíteni.

² Ezenkívül a jó proxy-nak ki kell szűrnie minden korrelációt a nem megfigyelhető közvetett változó és a regressziós egyenletben levő többi magyarázó változó között. Például az orvos-beteg kapcsolatok változékonyságát egyértelműen meghatározza az intézmény-váltások száma, ehhez az összefüggéshez nem ad hozzá a többi magyarázó változó.

4.4 Magyarázó változók a regresszióban

A modell építésekor néhány magyarázó változó szerepeltetésére különös figyelmet kell fordítani:

- Az eredményváltozó időbeli alakulását érdemes időbeli lineáris, kvadratikus vagy egyéb polinomiális trenddel megragadni. Ezek a trendek különböző csoportokra különbözők lehetnek, amit a csoportot definiáló dummy trenddel való interakciója segítségével ragadhatunk meg. Amennyiben az intézkedési változók keresztmetszetben is erősen szórnak, akkor akár évenkénti dummykat is használhatunk. (Ugyanakkor, ha van olyan intézkedés, amely csak időben szór, akkor az évenkénti dummy használata esetén az intézkedés hatását nem tudjuk azonosítani, és kevés év esetén már az időbeli trend használatával is probléma lehet.) Az éven belüli szezonális kiszűrésére érdemes havi vagy negyedéves dummykat alkalmazni a regresszióban.

- A különböző kórházi osztályok teljesítményének kontrollálására osztályszintű dummy alkalmazhatók a regresszióban. Figyeljünk ugyanakkor arra, hogy ha ezeket alkalmazzuk, akkor például megye-dummykat nem kell használni, hiszen az osztályszintű bontás részletesebb a megyeinél.

- Ha dummykat alkalmazunk, akkor legyen egy baseline csoport, amihez képest a többi csoport viselkedését vizsgáljuk. (Tehát a csoportok számánál eggyel kevesebb dummyt kell definiálnunk.)

- A magyarázó változók mellett (helyett) azok függvényei is szerepelhetnek a regresszióban. Az életkor hatását például sokszor kvadratikus függvényként szokták modellezni. Egy másik gyakori transzformáció a logaritmus: például az ágyszámváltozás helyett könnyebben interpretálható magyarázó változó a logaritmus változása (amely kis változások esetén közelítően a százalékos változással egyezik meg).

4.5 Modellforma és a paraméterek értelmezése

Egyszerű lineáris modell. A regresszió eredményváltozójától függően más és más modellformát kell választanunk. Folytonos változó esetén a legkézenfekvőbb a lineáris regressziós modell választása, ami a következőképpen néz ki. Legyen y az eredményváltozó, X_i ($i=1, \dots, k$) a magyarázó változók (beleértve az intézkedési változókat és a konstans is) és β_i ($i=1, \dots, k$) a becslendő paraméterek. Ekkor

$$(1) \quad y = \sum_{i=1}^k \beta_i X_i + u$$

ahol u a nem megfigyelt hiba.

Paraméterek értelmezése. A paraméterek értelmezése lineáris modellben egyszerű. Tegyük fel, hogy az X_1 magyarázó változó (speciálisan az intézkedési változó) dummy, és a hozzá tartozó paraméter β_1 . Ekkor az intézkedés az eredményváltozó várható értékét β_1 -gyel változtatja, függetlenül a többi magyarázó változótól:

$$(2) \quad E(y | X_1 = 1) - E(y | X_1 = 0) = \beta_1.$$

Ha folytonos a magyarázó változó, akkor a paraméter azt jelenti, hogy a változó (kis) egységnyi megváltozása mennyivel változtatja meg az eredményváltozót, függetlenül a többi magyarázó változótól.

A paraméterek egyszerű értelmezése mutatja a lineáris modell könnyű kezelhetőségét. Megjegyzendő ugyanakkor, hogy a lineáris regressziós keret nem minden folytonos eredményváltozó modellezésére megfelelő. Ha az eredményváltozó például időtartam-típusú, akkor ún. időtartam-moddellel (hazard-moddellel) jobban becsülhetők a hatások, míg cenzorált

eredményváltozó esetén cenzorált regressziós modellt kell alkalmazni. Ezekkel a speciális modellekkel itt nem foglalkozunk, az érdeklődő olvasó megtalálja őket a hivatkozott angol nyelvű ökonometriai irodalomban.

Logit és probit modell. Ha az eredményváltozó dummy (és ez a helyzet az egészségpolitikai intézkedések indikátor alapú hatásvizsgálatakor), akkor az eredményváltozó bekövetkezési valószínűségét kell modelleznünk. Általában logit vagy probit függvényformát szoktak választani. Mindkettőben

$$(3) \quad P(y = 1) = G\left(\sum_{i=1}^k \beta_i X_i\right),$$

ahol P jelöli a valószínűséget, és G egy eloszlásfüggvény (azaz a valós számokat a $(0,1)$ intervallumra leképező függvény). Speciálisan a probit modell esetében $G = \Phi$, azaz a standard normális eloszlásfüggvény, a logit modell esetén pedig G az ún. logisztikus függvény. Alább a logit specifikációt fejtjük ki részletesebben, mert azt az orvosi irodalom gyakrabban használja. Ekkor a valószínűség tehát a következő alakú:

$$(4) \quad P(y = 1) = \frac{e^{\sum_{i=1}^k \beta_i X_i}}{1 + e^{\sum_{i=1}^k \beta_i X_i}},$$

vagy ezzel ekvivalens felírásban:

$$(5) \quad \log \frac{P(y = 1)}{1 - P(y = 1)} = \sum_{i=1}^k \beta_i X_i$$

A $P/(1-P)$ arányt odds-nak nevezzük. A logit modell tehát a log odds és a magyarázó változók között lineáris kapcsolatot feltételez.

Paraméterek értelmezése. Probit vagy logit modellben bonyolultabb a helyzet, mint egyszerű lineáris modellben, ugyanis a valószínűség százalékpontos megváltozása egy dummy magyarázó változó hatására függ a többi magyarázó változó aktuális értékétől is:

$$(6) \quad P(y = 1 | X_1 = 1) - P(y = 1 | X_1 = 0) = G\left(\beta_1 + \sum_{i=2}^k \beta_i X_i\right) - G\left(\sum_{i=2}^k \beta_i X_i\right).$$

A gyakorlatban sokszor a többi magyarázó változó átlagos értékénél szokták kiértékelni a hatást ("parciális hatás az átlagnál"), vagy minden megfigyelésre kiszámítják a hatást és azokat átlagolják ("átlagos parciális hatás").

Logit specifikáció esetén az (5) egyenlet alapján van egyszerűbb értelmezése is a paramétereknek. Ha egy dummy változó (pl. intézkedés) paramétere β , akkor az eredmény bekövetkezési valószínűségének odds-át az intézkedés kb. $100\beta\%$ -kal (pontosan $(e^\beta - 1)$ -szeresére) változtatta. Például ha intézkedés nélkül a bekövetkezési valószínűség (a többi magyarázó változó adott értéke mellett) 20% (azaz $\text{Odds} = 0,2 / (1 - 0,2) = 0,25$) és $\beta = 0,1$, akkor intézkedés után (minden egyéb változatlansága mellett) az odds kb. $0,25 * 1,1 = 0,275$, és így a bekövetkezési valószínűség kb. $0,275 / (1 + 0,275) = 0,216$ -ra, tehát 1,6 százalékponttal növekedett. Ha viszont a kezdeti bekövetkezési valószínűség 2%, akkor a növekmény csak kb. 0,2 százalékpont. Megjegyzendő ugyanakkor, hogy ha a bekövetkezési valószínűségek kicsik, akkor az odds százalékos változása közel van P százalékos változásához, tehát az intézkedés kb. $100\beta\%$ -kal (de továbbra sem százalékponttal!) változtatja meg a bekövetkezési valószínűségeket. Például a pszichiátriai gyakorlatban a 30 napon belül visszavett betegek

aránya átlagosan csupán 2%, tehát $\beta=0,1$ esetén mondhatjuk azt, hogy az intézkedés ezt az arányt kb. 10%-kal emelte (azaz például az átlagnál 2%-ról 2,2%-ra).

Ha egy magyarázó változó folytonos, akkor hasonló a helyzet a fentiekhez. A logit keretben azt jelenti a paraméter, hogy a változó (kis) egységnyi megváltozása az eredményváltozó odds-át kb. $100\beta\%$ -kal változtatja meg. Itt is kiszámolhatjuk a konkrét százalékpontos változást a magyarázó változók átlagértékénél, vagy minden megfigyelésre kiszámolhatjuk a hatást és azokat átlagolhatjuk.

Ha az intézkedés folytonos változót érint, akkor azt is figyelembe kell vennünk, hogy intézkedés nélkül mennyi lett volna a magyarázó változó értéke (például az egyes megyékben mennyivel változott volna intézkedés nélkül az ágyszám). Az intézkedés hatásának meghatározásakor minden megfigyelésre kiszámítjuk a modell által előrejelzett eredményváltozókat intézkedés nélkül és intézkedéssel, és ezek különbségét átlagolva megkapjuk az intézkedés átlagos parciális hatását.

4.6 Paraméterek becslése és tesztelése, modellszelekció, illeszkedésvizsgálat, diagnosztika

Becslés és a paraméterek tesztelése. A modellek (egyszerű lineáris, logit, probit, időtartam) becslését a szokásos statisztikai programcsomagok automatikusan elvégzik a szükséges technikával (legkisebb négyzetek, maximum likelihood stb.), így a becslési eljárásra nem térünk ki részletesen. Megemlítjük viszont, hogy egyszerű lineáris modell esetén mindig érdemes a becslést heteroszkedaszticitásra robusztus (ún. White-féle) standard hibákkal elvégezni, ami ma már a legtöbb programcsomagban beépített opció. (Az érdeklődők az ökonometriai technikák részletes leírását megtalálhatják Wooldridge (2002) tankönyvében, illetve intuitív kifejtését Wooldridge (2009) könyvében.)

A becslés outputja tartalmazza minden parameter pontbecslését és a becslés standard hibáját. A kettő hányadosaként kapjuk a t-statisztikát, amely meghatározza a paraméterek egyenkénti szignifikanciáját: ha a t-statisztika abszolút értéke nagyobb 1,96-nál, akkor a paraméter 5%-on, ha pedig 2,58-nál, akkor 1%-on szignifikáns.³ A legtöbb programcsomag kiírja a p-értéket is: minél kisebb a p-érték, annál szignifikánsabb a paraméter. Szokásosan akkor tekintünk egy változót szignifikánsnak, ha a p-érték kisebb 5%-nál.

Több paraméter együttes tesztelése is elvégezhető a Wald-, LM- és likelihood arány (LR) tesztek valamelyikével. Sok programcsomagban ez automatikus opció, de amennyiben nem, akkor az LR-teszt kézzel is könnyen kiszámolható. Megbecsüljük az általános modellt és azt a modellt is, ahol a tesztelendő paraméterek nullára vannak beállítva, és a két model log-likelihoodja különbségének kétszerese lesz a tesztstatisztika. (A modell log-likelihood értékét mindig tartalmazza a becslési output.) Ezt a tesztstatisztika értéket a tesztelendő paraméterek számával megegyező szabadsági fokú χ^2 -eloszlás 5%-os kritikus értékével kell összehasonlítani.

Modellválasztás. A "legjobb" model kiválasztása inkább művészet, mint egzakt tudomány. A modellezés során kiindulhatunk úgy, hogy kezdetben minden, potenciálisan fontos változót szerepeltetünk, majd azt fokozatosan szűkítjük a nem szignifikáns változók vagy nem szignifikáns változócsoporthoz kihagyásával. Ha egy változócsoporthoz (például az összes megye dummy) együtt szignifikáns, de külön-külön sok a nem szignifikáns változó, akkor érdemes értelemeszerűen összevonni őket, például egyes megyecsoporthoz régiós dummykat

³ Ezek az értékek nagy mintában érvényesek, de a gyakorlatban elérhető egészségügyi mikroadatok többnyire ilyenek.

kialakítani. Hasonlóan az intézményi dummyk részösszevonására is szükség lehet. Egy másik modellezési filozófia szerint viszont érdemes egy olyan szűkebb modellből kiindulni, ami csak az eredményeket biztosan befolyásoló és az intézkedési változóval összefüggésben levő változókat tartalmazza, és ezt a modellt fokozatosan bővíteni – ha szükséges – további szignifikáns változókkal.

Összességében a végső modellben inkább szignifikáns változók (vagy legalábbis együttesen szignifikáns változócsoporthoz tartozók) legyenek, bár benthagyható néhány olyan nem szignifikáns változó is, ami a várt hatásokat megalapozó (közgazdasági vagy orvosi) elmélet és az interpretáció szempontjából fontos. Még ezzel a feltétellel is sok "jó" model lehetőség, hiszen nem egymásba ágyazott modellek esetén (azaz amikor az egyik modell változóhalmaza sem részhalmaza a másiknak) formális szignifikanciavizsgálat nem lehetséges. Ekkor különböző modelszelekciós kritériumok (pl. Akaike vagy Bayes-i információs kritérium) használhatók. Végső soron, ha több modell is "jó" tűnik, akkor érdemes megvizsgálni, hogy az intézkedések hatásai szempontjából hasonló következtetésekre jutnak-e, azaz megállapításaik elég robusztusak-e.

Modellilleszkedés ("goodness of fit") vizsgálata. A lineáris regressziós keretben a leggyakrabban használt modellilleszkedési mutató az R^2 , amely azt mutatja meg, hogy a modell milyen arányban magyarázza meg az eredményváltozó varianciáját (így tehát 0 és 1 közé esik). Logit és probit modellekben egy ehhez hasonló mutatót, a pszeudo R^2 -et,⁴ vagy egyszerűbb mérőszámként a jól klasszifikált esetek százalékos arányát használják. Ez utóbbi kiszámításakor minden egyes megfigyelésre meghatározzák a modell által (3) alapján előrejelzett valószínűséget, és úgy tekintik, hogy a modell az eredményváltozó 1 értékét jelzi előre, ha ez a valószínűség 0,5-nél nagyobb, egyébként pedig 0-t. Végül megszámlálják, hogy így a modell hány százalékban találja el a tényleges kimenetelt. Ez a mutató félrevezető lehet, ha nagyon alacsony az egyik eset (0 vagy 1) bekövetkezési valószínűsége, ezért küszöbértéknek valamikor nem 0,5-öt, hanem például az eredményváltozó átlagos bekövetkezési arányát tekintik.

Modelldiagnosztika. Az egyszerű lineáris modellben a reziduálisok (hibatagok) vizsgálatának van nagy jelentősége. Ha jó a modellspecifikáció, akkor a reziduálisoknak a magyarázó változókra vett feltételes várható értéke zérus kell, hogy legyen. Így a hibatagok és a magyarázó változók között az ábrázolás során kirajzolódó *függvényyszerű kapcsolat* arra utal, hogy a magyarázó változó függvényformája nem jól specifikált (például lineáris helyett kvadratikus az összefüggés, és ennek fényében meg kell változtatni a modellt). Ha a reziduálisok szórása függ valamely magyarázó változótól, akkor *heteroszkedaszticitásról* beszélünk, ennek jelenléte esetén a becslés során mindenképpen szükséges a heteroszkedaszticitásra robusztus standard hibák használata (ami a legtöbb programcsomagban beépített opció). A reziduálisok ezenkívül időben vagy térben korreláltak lehetnek, ekkor *autokorrelációról* beszélünk. Ez inkább panel- vagy idősoros adatok esetén gyakori probléma, jelenléte esetén a becslést autokorrelációra robusztus standard hibákkal (pl. Newey-West-féle eljárással) kell elvégezni, vagy kvázi-differenciálással az autokorrelációt megszüntetni. Gyakran a reziduálisok *normális eloszlását* is ellenőrzik, ennek azonban egyszerű lineáris modell esetén kevés jelentősége van, a paraméterbecslések ugyanis ilyenkor nagy mintában a normális eloszlás feltételezése nélkül is kedvező tulajdonságúak (azaz konzisztensek és aszimptotikusan normálisak).

Diszkrét függő változós modellben a fenti diagnosztikai problémák (pl. heteroszkedaszticitás) a valószínűséget meghatározó G függvény alakjának téves specifikációjában csapódnak le. A diszkrét változós modellek formális diagnosztikai vizsgálata bonyolultabb az egyszerű lineáris

⁴ Definíció: $1 - L_{ur} / L_0$, ahol L_{ur} a modell log-likelihoodját, L_0 pedig a csak a konstanst tartalmazó modell log-likelihoodját jelöli.

modellénél, ezért az alkalmazás-orientált tanulmányok általában azt az utat követik, hogy többféle specifikációt (pl. logit és probit) is becsülnek, és ha az átlagos parciális hatások nem térnek el jelentősen egymástól, akkor megbíznak a modellek eredményében.⁵

5. Eredmények és a korlátok összegzése

Statisztikai és "lényegi" szignifikancia. Az intézkedések hatásának fenti módon kiszámolt pontbecslése és értelmezése mellé mindig oda kell tenni a szignifikanciaszintet, valamint vizsgálni kell a hatás közgazdasági ("lényegi") szignifikanciáját is. Nagy mintában gyakran előfordul, hogy a változók mindegyike erősen szignifikáns, de csak néhány változónak van lényegi hatása. A hatások nagyságrendjével tehát mindig tisztában kell lennünk.

Endogenitás – exogenitás. Mint korábban már említettük, fontos hangsúlyozni az eredmények értékelésekor, hogy a becsült paramétereket csak akkor lehet tényleges *ok-okozati összefüggésként* értelmezni, ha a szereplő magyarázó változók exogének, azaz minden egyéb releváns tényezőt kiszűrtünk a modellben. Az exogenitás több – egymással összefüggő – okból csorbulhat, azaz több okból léphet fel endogenitás.

Az egyik eset, amikor olyan *nem megfigyelhető változók* vannak a háttérben, amelyek mind a magyarázó változókra, mind az eredményváltozóra hatással vannak. Ezeket – éppen mivel nem megfigyelhetők – nem tudjuk további magyarázó változóként szerepeltetni. Például ha a 30 napon belüli kórházi visszavételeket vizsgáljuk, és magyarázó változóként az aktív ápolási időt szerepeltetjük, akkor ennek együtthatója (még betegségcsoportra és egyéb változókra való kontrollálás után) sem fogja megadni az ápolási idő pontos *hatását* (ok-okozati összefüggését) a visszavételre. Az azonos betegségcsoporton belüli súlyosabb eseteket ugyanis valószínűleg később engedik haza és nagyobb eséllyel veszik vissza 30 napon belül, viszont a "súlyosságot" nem lehet pontosan mérni. Ennek a problémának a kezelésére szolgál például az instrumentális változók módszere, amikor egy olyan exogén "sokkot" keresünk, amely az eredményváltozóra csak az adott magyarázó változón keresztül hat (részletesebben ld. a Függelékben illetve Wooldridge 2002, 2009 tankönyveiben). A gyakorlatban a nem megfigyelhető változó problémára mindig gondolni kell az elemzés során.

Végül, egy másik, az egészségügyi alkalmazásokban fontos eset a *szelekciós probléma*. Ha például egy betegségből való felépülés időtartamát akarjuk vizsgálni, csak olyanokról lesz adatunk, akik elmentek az orvoshoz és diagnosztizálták a betegségüket – akik enélkül gyógyultak meg, sosem kerülnek be a mintába. A szelekció nem véletlenszerű: a fenti esetben például összefügghet a tünetekkel: akiknek enyhébbek a tünetei, azok kisebb eséllyel mennek orvoshoz. A szelektált minta ilyen esetben nem reprezentálja a teljes érintett kört, így torz eredményhez vezethet, ha erre korlátozzuk az elemzést. A megoldás első lépése a szelekciós lehetőségek végiggondolása, azonosítása, a következő pedig, ahol lehetséges, a szelekciós folyamat modellezése – bővebben lásd a megadott ökonometriai tankönyveken túl pl. Kézdi (2005) cikkét. A szelekciós torzításra is mindig gondolni kell az alkalmazások során.

⁵ Hangsúlyozandó, hogy a β paraméterek lehetnek nagyon eltérőek a különböző specifikációkban. A fontos az, hogy az átlagos parciális hatások hasonlóak legyenek.

6. Hivatkozások

Grun, R. (2006): Monitoring and evaluating projects, a step-by-step primer on monitoring, benchmarking and impact evaluation. HNP Discussion Paper, World Bank.

Khandker, S. R., Koolwal, G. B., Samad, H. A. (2010): Handbook on impact evaluation, quantitative methods and practices. World Bank.

Kézdi G. (2005): James M. Heckman. In: Közgazdasági Nobel-díjasok 1969-2004 (szerk.: Bekker Zs.). KJK-Kerszöv. http://www.personal.ceu.hu/staff/Gabor_Kezdi/Publications/Kezdi-2005-NobelKotet-Heckman.pdf

Ramanathan, R. (2003): Bevezetés az ökonometriába alkalmazásokkal. Panem.

Wooldridge, J. M. (2002): Econometrics of cross section and panel data. The MIT Press.

Wooldridge, J. M. (2009): Introductory econometrics, a modern approach. South Western Publishing.

7. Függelék: Releváns becslési módszerek⁶

7.1. Tervezett kísérletek

A módszer leírása. A tervezett kísérletekben a program potenciális résztvevőit (egyének, intézmények, térségek) véletlen kiválasztással két csoportba osztják. A „kezelt csoport” tagjai részt vehetnek a programban, a „kontroll csoport” tagjai nem. Ideális esetben így a két csoport összetétele azonos, ezért a kontroll csoport eredményváltozó-eloszlása jól írja le a kezelt csoport tényellentétes eredményeit, vagyis azt, hogy milyen eredményeket értek volna ők el, ha nem vesznek részt a programban. A kezelt és a kontroll csoport átlagos eredményeinek összehasonlítása ezért megfelelő becslése a program átlagos hatásnak.

A módszer legfontosabb elemei:

a potenciális résztvevők azonosítása

véletlenszerű kiválasztás kezelt és kontroll csoportokba

adatfelvételek

a véletlenszerűség ellenőrzése (a csoportok összehasonlítása háttérváltozók szerint)

az eredményváltozók átlagainak összehasonlítása

Előnyök és hátrányok. A tervezett kísérlet a természettudományok klasszikus módszere. Elméleti szempontból a legtisztább és leginkább hihető eredményt produkáló hatásvizsgálati módszerek. Pontosságuk mellett fontos előnyük egyszerűségük. Hátrányuk, hogy társadalompolitikai alkalmazhatóságuk korlátozott. A tervezett kísérleteket a beavatkozásokba be kell építeni azok megtervezésekor; utólagos elemzések esetén a módszer nem alkalmazható. Az embereken való kísérletezés pedig nyilvánvaló morális és politikai problémákat jelent, és a társadalmi kísérletek lebonyolítását egy sor praktikus akadály is nehezíti.

⁶ A módszerek leírása kisebb változtatásokkal Kézdi Gábor összefoglalója, amelyet az IFUA által vezetett értékelési konzorcium Nemzeti Fejlesztési Ügynökségnek benyújtott ajánlatából vettünk át.

7.2. Diszkontinuitás modellek

A módszer leírása. A diszkontinuitás (regression discontinuity design) modellek egy nem kísérleti elemzési módszert jelölnek. Akkor alkalmazhatók, ha a programba kerülés valamilyen folytonos változóhoz köthető, melynek egy meghatározott küszöbértéke dönti el, hogy az illető részesülhet-e a "kezelésben" vagy nem. A leggyakoribb példák a korhatárhoz kötött programok (pl. 3 év alattiak vehetnek csak részt). A küszöbérték egyik oldalán a részvétel valószínűsége 0, a másik oldalán pozitív. A diszkontinuitás modellek alapvető ötlete az, hogy a határ egyik oldalán található egyének (egy hónap híján 3 évesek) és a határ másik oldalán található egyének (éppen betöltötték a 3. évet) valószínűleg nagyjából ugyanolyan összetételűek minden szempontból. Formálisan: a küszöbértékben folytonos minden tulajdonság, így az eredményváltozó is – feltéve hogy a program nem létezik. A kezelés valószínűsége ezzel szemben diszkontinuos (szakadást mutat) a küszöbértékben. A két csoport eredményváltozói így csak akkor különbözhetnek egymástól, ha a programnak van hatása (hiszen a két csoport gyakorlatilag ugyanolyan minden egyéb tekintetben).

A módszer legfontosabb elemei:

- a szakadási pont azonosítása a beavatkozás sajátosságai alapján adatfelvételek
- a szakadási pont empirikus azonosítása (a részvételi valószínűségek meghatározása); a folytonosság ellenőrzése azon változók esetében, amely nem köthető a beavatkozáshoz
- a diszkontinuitási minta kialakítása (a határ két oldalán milyen távolságban húzzuk meg az elemzésbe bevontak körét)
- a hatások becslése az eredményváltozók összehasonlítása alapján (átlagok összehasonlítása; parametrikus lineáris vagy polinomiális regressziók; lokálisan simított nemparametrikus lowess regressziók stb.)

Előnyök és hátrányok. Ideális esetben – azaz, amennyiben megfelelő mennyiségű és minőségű adat áll rendelkezésre, megfelelő mértékű a szakadás a beavatkozási valószínűségben, valamint a programhoz nem köthető változók folytonosak - a diszkontinuitás alapú hatásvizsgálatok megbízhatóan becslik a program hatását (legalábbis a diszkontinuitási pont körül). Legfontosabb hátránya, hogy csak olyan esetekben alkalmazható, ahol a beavatkozás valószínűségének van szakadása. A programok tervezésénél kell ilyen szakadást beépíteni (pl. scoring rendszer), ugyanakkor utólagos hatásvizsgálatnál ez már nem lehetséges. További hátránya a jelentős adatigény, elsősorban a mintanagyság tekintetében, hiszen sok megfigyelésre van szükség a határ két oldalán levő szűk területen.

7.3. Párosítási modellek és egyéb propensity score-alapú módszerek

A módszer leírása. A párosítási („matching”) és egyéb propensity score-alapú modellek olyan elemzési módszerek, amelyeket elsősorban statisztikai hatásvizsgálatok során alkalmaznak, nem kísérleti szituációban. A párosítási modellek lényege az, hogy a programban részt vevő valamennyi egyént párosítja egy vagy több nem résztvevővel. A módszer célja az, hogy a létrehozott párok eredményváltozóinak az összehasonlításával minden egyes résztvevőre megbecsülje a program hatását, vagyis a beavatkozás eredményeként elért eredmény és a tényellentétes (a beavatkozás hiányában elért) eredmény különbségét. A párok tagjai közötti különbség akkor tudja pontosan megragadni a program-hatást, ha a párok minden olyan

változó tekintetében hasonlóak, amelyek befolyásolják az eredményt akármelyik (kezelt vagy nem kezelt) állapotban.

A párosítás általában egy úgynevezett propensity score alapján történik. A propensity score a programban való részvétel becsült valószínűsége, amit a tényleges résztvevőkre és nem résztvevőkre is meg kell becsülni. A résztvevőket hozzájuk hasonló propensity score-al rendelkező nem résztvevőkkel párosítjuk. A párosítás lehet egy az egyhez, egy a többhöz, de akár több az egyhez vagy több a többhöz is. A párosított mintában sokféle módszerrel lehet az eredményváltozókat összehasonlítani: a különbségek egyszerű átlaga mellett a különféle kernel súlyozású átlagokat alkalmazzák a leggyakrabban. Gyakori megoldás, hogy az elemzési minta kialakítása párosítási modell segítségével történik, maga az elemzés pedig más módszereket alkalmaz. A propensity score a párosítás mellett más módon is alkalmazható a hatások becslésére, akár kontrol függvényekben, akár közvetlen módon, a minták átsúlyozása útján.

A módszer legfontosabb elemei:

adatfelvételek

a párosított minta kialakítása (propensity score becslés és ez alapján a párosítás elvégzése)

a párosított mintán az eredmények összehasonlítása, és ezek alapján a hatás becslése (egyszerű és súlyozott átlagolások alapján, pl. kernel súlyok, vagy bonyolultabb regressziós módszerek alapján)

Előnyök és hátrányok. A párosítási modellek nagy előnye, hogy utólagos hatásvizsgálatok esetében a program designtól függetlenül alkalmazhatók. Megfelelő mennyiségű és minőségű adatok esetén, és amikor a tervezett kísérletek vagy a diszkontinuitás modellek nem alkalmazhatók, a párosított mintáknál általában nincs jobb minta-kialakítási módszer hatásvizsgálatokhoz. További előnye, hogy rugalmasan kombinálható regressziós módszerekkel. Hátránya a nagy adatigény, nemcsak mintanagyság, hanem a megfigyelt változók mennyiségének tekintetében is.

7.4. Regressziós modellek, kontroll függvények, panel és difference-in-differences modellek

A módszer leírása. A regressziós modellek elemzési módszert jelölnek, amelyeket elsősorban statisztikai hatásvizsgálatok során, nem pedig kísérleti szituációban alkalmaznak. A regressziós modellek közé tartoznak mindazon módszerek, amelyeket egyenlet(ek) formájában lehet felírni, ahol az eredmény az egyenlet(ek) bal oldalán, a beavatkozás ténye (a részvétel) az egyenlet(ek) jobb oldalán szerepel. Az átlagok egyszerű összehasonlítása is felírható regressziós egyenlet formájában, a regressziókat azonban általában akkor használják, ha kontrollálni akarnak egyéb változók hatásaira (ki akarják szűrni a program látszólagos hatásából azokat a különbségeket, amelyek nem a program hatása miatt figyelhetők meg, hanem mert a résztvevők szisztematikusan mások, mint a nem résztvevők).

Kontroll függvények esetében a kontroll változók szerepeltetésének a formáját a hatásvizsgálat diktálja (polinomok, interakciók a részvétel dummyval). A különbségek különbsége (difference-in-differences, röviden: diff-in-diffs) modellek esetében a kérdés nem az eredményváltozó szintje, hanem annak változása a program előtti értékhez képest. Szűk értelemben a diff-in-diffs modellek az eredményváltozó változását vizsgálják. Tágabb értelemben a diff-in-diffs azt jelenti, hogy a kontroll változók között szerepel az eredményváltozó a program előtti értéke is. A panel modellek tekinthetők a diff-in-diffs modellek további általánosításának, ahol nemcsak az eredmények, hanem a többi változó

hatását is változásokban specifikáltak (vagy a saját átlagtól való eltérésként, pl. fixed-effects). Lineáris regressziók esetében a függvények lineárisak (a paraméterekben), nemlineáris regressziók esetében nem (pl. probit vagy logit valószínűségi modellek, tobit cenzorálási modellek, vagy időtartam – más néven hazard – modellek).

A módszer legfontosabb elemei:

az elemzési minta kialakítása

adatfelvételek

a becült modellek specifikációja (a regresszió formája pl. probit, logit, exponenciális hazard; a bevonandó kontroll változók köre; a kontrollfüggvény formája, stb.)

becslés (OLS, GLS, SUR, nonlinear least squared, Maximum Likelihood, stb.)

Előnyök és hátrányok. A regressziós modelleket elvileg a párosítási modellek alternatívájaként is tekinthetjük (ideális esetben azonos eredményt kell mutatniuk), a gyakorlatban a két módszert inkább egymás kiegészítőiként alkalmazzák. A regressziók alkalmazásának feltételei azonban szigorúbbak, ezért érdemes kombinálni őket a párosított minták módszerével. Megfelelő kombináció esetén előnyei és hátrányai megegyeznek a párosítási módszerek előnyeivel és hátrányaival: utólagos hatásvizsgálatok esetében a program felépítésétől függetlenül alkalmazhatók, de sok adatot igényelnek, nemcsak mintanagyság, hanem a megfigyelt változók mennyiségének tekintetében is.

7.5. Instrumentális változós módszerek, szelekciós modellek, switching regressziók

A módszer leírása. Az instrumentális, szelekciós és switching modellek elemzési módszert jelölnek, amelyeket elsősorban statisztikai hatásvizsgálatok során alkalmaznak, nem kísérleti szituációban. A regressziós modellek alternatívái. Alapjában regressziós modellekről van szó, a programban való nem véletlenszerű („endogén”) részvétel miatt. Legfontosabb jellemzőjük egy olyan mechanizmus kihasználása, amely befolyásolja a programban való részvételt, de az eredményekre önmagában nincs hatása. Ezt a mechanizmust ragadja meg az úgynevezett instrumentális változó. A szelekciós és switching modellek tekinthetők az instrumentális modellek speciális eseteinek, ahol a mechanizmusokat leíró egyenletek nem lineárisak.

A módszer legfontosabb elemei:

az elemzési minta kialakítása

adatfelvételek

a becült modellek specifikációja (a regresszió formája pl. probit, logit, exponenciális hazard; a bevonandó kontroll változók köre; a kontrollfüggvény formája, stb.)

az instrumentum megtalálása

becslés (OLS, GLS, SUR, nonlinear least squared, Maximum Likelihood, stb.)

Előnyök és hátrányok. Az instrumentális, szelekciós és switching modellek nagy előnye, hogy direkt módon kezelik a programban való részvétel endogenitását. Hasonlóan a regressziós modellekhez, utólagos hatásvizsgálatok esetében a program designától függetlenül alkalmazhatók, de sok adatot igényelnek, nemcsak mintanagyság, hanem a megfigyelt változók mennyiségének tekintetében is. Legnagyobb hátrányuk, hogy csak érvényes instrumentum esetén vezetnek az egyszerűbb regressziós modelleknél megbízhatóbb eredményekre, ilyen instrumentumokat azonban ritkán lehet találni.