MAGYAR TUDOMÁNYOS AKADÉMIA
KÖZGAZDASÁGTUDOMÁNYI KUTATÓKÖZPONT

BUDAPESTI MUNKAGAZDASÁGTANI FÜZETEK

BWP. 2004/2

Az aktív foglalkoztatáspolitikai programok hatásvizsgálatának módszertani kérdései

KÉZDI GÁBOR

Magyar Tudományos Akadémia Közgazdaságtudományi Kutatóközpont
Munkaerőpiaci Kutatások Műhelye

Budapesti Közgazdaságtudományi és Államigazgatási Egyetem
Emberi Erőforrások Tanszék

Budapest

Az aktív foglalkoztatáspolitikai programok hatásvizsgálatának módszertani kérdései

KÉZDI GÁBOR

Budapesti Munkagazdaságtani Füzetek

BWP. 2004/2

2004. január

Budapesti Munkagazdaságtani Füzetek 2004/2. szám

Magyar Tudományos Akadémia Közgazdaságtudományi Kutatóközpont,
Munkaerőpiaci Kutatások Műhelye
Budapesti Közgazdaságtudományi és Államigazgatási Egyetem Emberi Erőforrások Tanszék

Az aktív foglalkoztatáspolitikai programok hatásvizsgálatának módszertani kérdései

Szerző: KÉZDI Gábor tudományos munkatárs, MTA Közgazdaságtudományi Kutatóközpont; egyetemi adjunktus, Budapesti Közgazdaságtudományi és Államigazgatási Egyetem, Közép-Európai Egyetem
E-mail: kezdi@econ.core.hu

A tanulmány "A foglalkoztatáspolitikai eszközök hatékonyságának értékelése" c. MTA – FMM5 kutatási téma keretében készült.

ISSN 1785-3788
ISBN 963 9321 98 2

Kiadja az MTA Közgazdaságtudományi Kutatóközpont
a "Közösen a jövő munkahelyeiért" Alapítvány és a
Magyar Közgazdász Alapítvány támogatásával

Budapest, 2004

BUDAPEST WORKING PAPERS ON THE LABOUR MARKET | **BUDAPESTI MUNKAGAZDASÁGTANI FÜZETEK**

BWP. 2000/1	Péter Galasi and Gyula Nagy	Are children being left behind in the transition in Hungary?
BWP 2000/2	Árpád Ábrahám and Gábor Kézdi	Long-run trends in earnings and employment in Hungary, 1972–1996
BWP 2000/3	Kőrösi Gábor	A vállalatok munkaerő-kereslete
BWP 2000/4	Kertesi Gábor	A cigány foglalkoztatás leépülése és szerkezeti átalakulása 1984 és 1994 között
BWP 2000/5	Fazekas Károly	A külföldi működőtőke-beáramlás hatása a munkaerőpiac regionális különbségeire Magyarországon
BWP 2000/6	Kertesi Gábor	Ingázás a falusi Magyarországon
BWP 2000/7	G. Kertesi–J. Köllő	Wage Inequality in East-Central Europe
BWP 2000/8	Károly Fazekas	The impact of foreign direct investment inflows on regional labour markets in Hungary
BWP 2000/9	Nagy Gyula	A munkanélküli-segélyezés Magyarországon a kilencvenes években
BWP 2001/1	János Köllő	The patterns of non-employment in Hungary's least developed regions
BWP 2001/2	Köllő János	A munkanélküli segélyrendszer 2000. évi szigorításának politikai támogatottsága
BWP 2001/3	Kertesi G. – Köllő J.	Ágazati bérkülönbségek Magyarországon
BWP 2001/4	Gábor Kertesi – János Köllő	Economic transformation and the revaluation of human capital – Hungary, 1986–1999
BWP 2001/5	Galasi P. – Nagy Gy.	Járadékjogosultság és elhelyezkedési esélyek
BWP 2001/6	Kertesi Gábor – Köllő János	A gazdasági átalakulás két szakasza és az emberi tőke átértékelődése Magyarországon
BWP 2001/7	Köllő János	A járadékos munkanélküliek álláskilátásai 1994 és 2001 tavaszán
BWP 2001/8	Galasi Péter– Nagy Gyula	A munkanélküli ellátás változásainak hatása a munkanélküliek segélyezésére és elhelyezkedésére
BWP 2001/9	Fazekas Károly	Az aktív korú állástalanok rendszeres szociális segélyezésével és közcélú foglalkoztatásával kapcsolatos önkormányzati tapasztalatok
BWP 2001/10	Júlia Varga	Earnings Expectations and Higher Education Enrolment Decisions in Hungary
BWP 2001/11	Köllő János	Meddig tart a rendszerváltás?

BUDAPEST WORKING PAPERS ON THE LABOUR MARKET | **BUDAPESTI MUNKAGAZDASÁGTANI FÜZETEK**

BWP 2002/1	Péter Galasi–Júlia Varga	Does Private and Cost-Priced Higher Education: Produce Poor Quality?
BWP 2002/2	Köllő János	Az ingázási költségek szerepe a regionális munkanélküli különbségek fenntartásában – Becslési kísérletek
BWP 2002/3	Gábor Kézdi	Two Phases of Labor Market Transition in Hungary: Inter-Sectoral Reallocation and Skill-Biased Technological Change
BWP 2002/4	Gábor Kőrösi	Labour Adjustment and Efficiency in Hungary
BWP 2002/5	Gábor Kertesi and János Köllő	Labour Demand with Heterogeneous Labour Inputs after the Transition in Hungary, 1992–1999 – and the Potential Consequences of the Increase of Minimum Wage in 2001 and 2002
BWP 2002/6	Fazekas Károly	A tartós munkanélküliek rendszeres szociális segélyezése és önkormányzati köz-foglalkoztatása Magyarországon, 2000-2001
BWP 2002/7	Zsombor Cseres-Gergely	Residential Mobility, Migration and Economic Incentives – the Case of Hungary in 1990–1999
BWP 2002/8	Kőrösi G.–Surányi É.	Munkahely-teremtés és -rombolás
BWP 2003/1	Ágnes Hárs	Channeled East-West labour migration in the frame of bilateral agreements
BWP 2003/2	Galasi Péter	Munkanélküliségi indikátorok és az állásnélküliek munkaerő-piaci kötődése
BWP 2003/3	Károly Fazekas	Effects of foreign direct investment on the performance of local labour markets – The case of Hungary
BWP 2003/4	Péter Galasi	Estimating wage equations for Hungarian higher-education graduates
BWP 2003/5	Péter Galasi	Job-training of Hungarian higher-education graduates
BWP 2003/6	Gábor Kertesi–János Kölő	The Employment Effects of Nearly Doubling the Minimum Wage – The Case of Hungary
BWP 2003/7	Nemes-Nagy József–Németh Nándor	A "hely" és a "fej". A regionális tagoltság tényezői az ezredforduló Magyarországán
BWP 2003/8	Júlia Varga	The Role of Labour Market Expectations and Admission Probabilities in Students' Application Decisions on Higher Education: the case of Hungary
BWP 2004/1	Gábor Kertesi	The Employment of the Roma – Evidence from Hungary

A Budapesti Munkagazdaságtani Füzetek a Magyar Tudományos Akadémia Közgazdaságtudományi Kutatóközpontjában működő Munkaerőpiaci Kutatások Műhelyének valamint a Budapesti Közgazdaságtudományi és Államigazgatási Egyetem Emberi Erőforrások Tanszékének közös kiadványa. A kiadványsorozat angol nyelvű füzetei "**Budapest Working Papers on the Labour Market**" címmel jelennek meg.

A sorozat egyes példányai a következő címeken szerezhetők be:

Szabó Irén, Budapesti Közgazdaságtudományi és Államigazgatási Egyetem Emberi Erőforrások Tanszék; 1093 Budapest, Fővám tér 8. Telefon/fax: 217-1936; E-mail: iszabo@workecon.bke.hu

Sándor Zsuzsa, MTA Közgazdaságtudományi Kutatóközpont Könyvtár; Budapest 1502 Pf. 26; Fax: 319-3136; E-mail: biblio@econ.core.hu

A kötetek letölthetők az MTA Közgazdaságtudományi Kutatóközpont honlapjáról:

http://www.econ.core.hu

KÉZDI GÁBOR

AZ AKTÍV FOGLALKOZTATÁSPOLITIKAI PROGRAMOK HATÁSVIZSGÁLATÁNAK MÓDSZERTANI KÉRDÉSEI

Összefoglalás

A tanulmány bemutatja a társadalompolitikai programok egyik legygyakrabban vizsgált típusa, az úgynevezett aktív foglalkoztatáspolitikai eszközök hatásvizsgálatának a legfontosabb módszertani problémáit, és azt, hogy ezekre a problémákra a szakma mai állása alapján milyen válaszok adhatók. Elsősorban a közvetlen hasznok mérésének problémáival foglalkozik. Azt a kérdést vizsgálja tehát, hogy az adott program milyen és mekkora hasznot hozott a benne részt vevők számára. Nem foglalkozik a közvetlen költségek mérhetőségének problémáival, és a közvetett hatások mérését külön fejezetben tárgyalja. A kérdések módszertani problémáit és a megoldási lehetőségeket általánosságban is vizsgálja, és "működés közben," egy-egy tipikus példán keresztül is bemutatja őket. A tanulmány bemutatja a közvetlen hatások mérésének általános problémáját és az ebből fakadó alapvető nehézségeket. Tárgyalja a mérési modellek típusait: a tervezett és természetes kísérleteket, a párosításon alapuló és az ökonometriai modelleket. Külön vizsgálja a kontroll csoport kiválasztásának problémáit nem kísérleti szituációkban. Bemutatja azokat a problémákat is, amiket a hatások mögötti mechanizmusok mérése jelent, azok lehetséges kezelésével együtt.

EVALUATING THE IMPACT OF ACTIVE LABOR MARKET PROGRAMS

BY GÁBOR KÉZDI

Abstract

The paper examines how to evaluate active labor market programs. It introduces the most important problems of measurement, and shows alternative solutions to those. Most of the paper focuses on the direct effects of such programs, that is the effect of the program on its participants. The paper does not discuss the measurement of direct costs. Measurement of the indirect effects is discussed in one separate chapter. The methodological problems are discussed in general and also through a few detailed case studies. Besides the major problems of identifying the direct effects, the paper examines in detail the different measurement strategies: controlled and natural experiments, matching models, and econometric models. Choice of the control group in non-experimental evaluations is crucial and is therefore discussed in a separate section. Measurement problems of the mechanisms underlying the effects is also examined.

Key words: active labor market program, program evaluation, identification problem
JEL Codes: J64, J68, C20

1. BEVEZETŐ

A modern jóléti államokban a társadalompolitikai programok legfontosabb mércéje az, hogy a társadalom számára milyen hasznokat hoznak és milyen költségekkel járnak. E költség-haszon elemzések rendkívül sok nehézségbe ütköznek minden oldalról. A közvetlen pénzügyi költségeket sem mindig egyszerű felmérni, igazán problematikus azonban azt megbecsülni, hogy a programokban résztvevők számára milyen és mekkora haszonnal jár az adott program. A közvetett hasznok és költségek – a nem résztvevőkre gyakorolt, a közgazdaságtanban általános egyensúlyinak is nevezett hatások – mérése általában még ennél is nehezebb feladat.

Jelen tanulmány célja az, hogy bemutassa a társadalompolitikai programok bizonyos típusa, az úgynevezett aktív foglalkoztatáspolitikai eszközök hatásvizsgálatának a legfontosabb módszertani problémáit, és azt, hogy ezekre a problémákra a szakma mai állása alapján milyen válaszok adhatók. Célunk az, hogy a lehető legegyszerűbben mutassuk be a problémákat és megoldási lehetőségeiket. A tanulmányban – mint a hatásvizsgálattal foglalkozó irodalom nagy részében is – elsősorban a közvetlen hasznok mérésének problémáival foglalkozunk. Azt a kérdést vizsgáljuk tehát, hogy az adott program milyen és mekkora hasznot hozott a résztvevők számára. Egyáltalán nem foglalkozunk a közvetlen költségek mérhetőségének problémáival, hiszen azokat programtól függően az azt adminisztráló szervezetek tudják legjobban mérni. A közvetett hatások mérését a tanulmány végén, külön tárgyaljuk. A kérdések módszertani problémáit és a megoldási lehetőségeket általánosságban is vizsgáljuk, és "működés közben," egy-egy tipikus vagy nagy hatású példán keresztül is bemutatjuk őket.

A társadalompolitikai programok hatásvizsgálatának (program evaluation) igen nagy, és folyamatosan növekvő módszertani irodalma van. Az első nagy hatású elemzés Ashenfelter (1978) cikke volt, a nyolcvanas évek második felétől kezdve (pl. Ashenfelter és Card, 1985) pedig igazi virágzásnak indult a módszertani kérdések vizsgálata. Több összefoglaló cikk is létezik (például Blundell és Costa Dias, 2002). A legátfogóbb, bár meglehetősen technikai és ezért nehezen követhető összefoglalás Heckman, LaLonde és Smith (1999) tanulmánya.

Az aktív foglalkoztatáspolitikai eszközök hatásvizsgálat szempontjából talán a legalaposabban vizsgált társadalompolitikai programok. Ennek valószínűleg az a legfontosabb oka, hogy viszonylag jól körülhatárolható intézkedések viszonylag jól körülhatárolható hatását szeretnénk mérni (szemben például a közoktatási vagy közegészségügyi programokkal, amelyek hatása

jóval hosszabb időn belül érvényesülhet, rengeteg más tényező hatásának közrejátszásával). Az aktív foglalkoztatáspolitikai eszközök munkanélküli, egyéb nem foglalkoztatott, vagy foglalkoztatott de alacsony keresetű (jellemzően hátrányos helyzetű) egyénekre fókuszálnak, és célul azok foglalkoztatási és/vagy kereseti helyzetének javítását tűzik ki. A leggyakoribb ilyen programok a következők:

- Képzési programok (pl. munkanélküliek átképzése)
- A munkaadók támogatása célszemélyek foglalkoztatása esetén
- A munkaadók támogatása már foglalkoztatott célszemélyek munkahelyi továbbképzése esetén
- Közcélú vagy egyéb non-profit foglalkoztatás
- Aktív segítség a munkakeresésben

Az aktív foglalkoztatási eszközök esetén a leggyakrabban vizsgált eredmény a foglalkoztatás, illetve az általa elérhető kereset. A vizsgált kérdés az, hogy milyen hatása van a résztvevők foglalkoztatására (keresetére) az adott program. A kérdés egész pontosan az, hogy *milyen eredményt értek el a résztvevők, összehasonlítva azzal, amit akkor értek volna el, ha nem vesznek részt a programban*. A megmérendő hatás tehát egy tényleges állapot (a program utáni eredmény) és egy úgynevezett *tényellentétes állapot* (milyen eredményt értek volna el a program hiányában) összehasonlításából adódik. Elvileg nyilvánvaló, hogy pusztán a program résztvevőinek vizsgálatával miért nem azonosítható a program hatása: a tényellentétes eredmények így nem mérhetők.

Úgy gondolhatnánk, hogy ha a gazdasági helyzet nem változott jelentősen a program futása alatt, akkor a tényellentétes mérés hiánya nem okozhat gyakorlati problémát. Ez azonban két okból sincs így. A kisebb probléma az, hogy az érintett csoportok iránti kereslet a gazdaság más folyamataitól eltérő módon változhat. Ez a probléma valójában mérési kérdés, és elvileg megoldható – bár gyakorlatilag sosem elhanyagolható. A másik probléma nem pusztán mérési kérdés. Ha a program résztvevőinek meg kell felelniük bizonyos kritériumoknak (pl. fél éve munkanélküli), akkor a veszélyeztetett (magas munkanélküliségi esélyű) csoportokon belül is azok vehetnek csak részt, akik éppen a szükséges időszakban voltak hátrányos helyzetben. Egy részük azonban tipikus esetben a program nélkül is kilábalna ebből, így csak a résztvevők pályáját elemezve ezt a potenciális önerős felemelkedést is a programnak tudnánk be. A jelenséget – amelyet Ashenfelter-horpadásnak (Ashenfelter's dip) is neveznek – és következményeit alább részletesen vizsgáljuk.

A tényellentétes eredmények mérése tehát elengedhetetlen a program hatásának azonosításához. A módszertani problémák pedig a tényellentétes eredmények mérésének nehézségéből fakadnak: valójában sosem tudhatjuk pontosan hogy mi történt volna a résztvevőkkel ha nem vesznek részt, hiszen egy egyén vagy részt vesz a programban, vagy nem. Ez a probléma, amit az ökonometriában *identifikációs problémának* neveznek, felfogható úgy, mint amely egy gondolatkísérlet kivitelezhetetlenségéből fakad. A gondolatkísérlet itt nem más, mint hogy egyes egyének eredményeit megmérjük úgy is hogy részt vettek a programban, és úgy is, hogy nem.

A gondolatkísérlet maga ugyan nem valósítható meg, a programban való részvétel véletlenszerű hozzárendelésével (tervezett vagy más néven kontrollált kísérletek révén) azonban az átlagos hatás identifikálható. Ha különböző emberekre különböző a hatás, akkor viszont ezek a különbségek kontrollált kísérletek esetében sem identifikálhatók teljes mértékben, csak mérhető ismérvek mentén. A véletlenszerű hozzárendelésnek azonban jelentős erkölcsi és politikai költségei vannak. Részben ezért, részben pedig valószínűleg a mérhetőség szempontjának figyelmen kívül hagyása miatt is ritkák az ilyen kísérletek. Bizonyos esetekben a hatásvizsgálattól független okokból kísérleti jellegű hozzárendelés történik: ezek az ún. természetes kísérletek. Az esetek nagy részében azonban sem tervezett, sem természetes kísérletek nem állnak a programban való részvétel hozzárendelése mögött. Ilyenkor valamilyen kontroll csoport választása az egyetlen járható – bár tökéletlen – út a hatások méréséhez. Megfelelő kontroll csoport kiválasztása rendkívül nehéz, ráadásul a nem jó kiválasztásból fakadó torzítások közvetlenül nem mérhetők.

A hatásvizsgálatok további kérdése az, hogy ha a programnak van kimutatható hatása, az milyen mechanizmusokon keresztül érvényesül. Ennek vizsgálata újabb problémákat vonhat maga után. Ha a programban való részvételnek pozitív hatása van (de nem feltétlenül tudjuk miért), akkor a programból kikerülők azonos egyéb tulajdonságok mellett sikeresebbek lesznek. Márpedig ha ezen egyéb tulajdonságok nem mind megfigyelhetők, de a sikert befolyásolják, akkor a sikeresek között a programban részt vevők és a kontroll csoport tagjai nem tekinthetők véletlen mintának még akkor sem, ha az eredeti kiválasztás véletlen volt. Ráadásul a programok általában jelentős időt vesznek igénybe, így a környezet jelentősen megváltozhat mire a részt vevők újra a munkaerőpiacon találják magukat, míg a kontroll csoport tagjai végig ott voltak. Ez az időbeli csúszás a program hatásának a mérését is nehezítheti (különösen nem kísérleti szituációkban), a mögötte húzódó mechanizmusok feltárását pedig tovább bonyolítja.

A tanulmány 2. része mutatja be részletesen a hatások mérésének általános problémáját és az ebből fakadó alapvető nehézségeket. A 3. rész bemutatja a mérési modellek típusait. A 4. rész tárgyalja a kontroll csoport kiválasztásának problémáját nem kísérleti szituációkban. Az 5. rész foglalkozik azokkal a problémákkal, amiket a mögöttes mechanizmusok mérése jelent, és bemutatja azok lehetséges kezelését is. A 6. részben röviden tárgyaljuk a közvetett hatások kérdéseit. Az utolsó rész foglalja össze a legfontosabb tanulságokat.[1]

2. A MÉRNI KÍVÁNT HATÁS MEGHATÁROZÁSA ÉS AZ IDENTIFIKÁCIÓS PROBLÉMA

Az alapvető mérési probléma bemutatásához egy leegyszerűsített szituációt tekintünk. Egyetlen eredményváltozó érdekel minket (pl. foglalkoztatás két év múlva), amit jól tudunk mérni. Feltesszük, hogy a résztvevők ugyanazt a "kezelést" kapják. Az egyszerűsítés kedvéért feltesszük azt is, hogy a program csak azokra hat, akik részt vesznek benne. Így csak a közvetlen (vagyis a résztvevőkre gyakorolt) hatásokkal kell foglalkoznunk.

Jelöljön i egy egyént, Y_i pedig az eredményt. Legyen D_i az a bináris változó, amely a részvételt jelöli: $D_i = 1$ ha i részt vesz a programban, $D_i = 0$ ha nem. A kísérleti statisztika nyelvén $D_i = 1$ ha az egyén kezelést kap (a kezelt csoportban van), 0 ha nem kap kezelést (a kontroll csoportban van). Jelölje Y_{1i} az i-ik egyén kezelés utáni eredményét, Y_{0i} pedig a kezelés hiányában bekövetkező eredményét.

Gondolatkísérletünkben minden egyes egyénre szeretnénk tudni a program hatását az eredményváltozóra:

$$\Delta_i = Y_{1i} - Y_{0i}$$

[1] A tanulmány módszertani jellegéből fakadóan meglehősen technikai. Végig igyekszünk azonban mindent a lehető legegyszerűbben tárgyalni. A fő problémák megértése a valószínűségszámítás alapjain (feltételes valószínűség, feltételes eloszlás, feltételes várható érték, stb.) túl különösebb felkészültséget nem igényel. A becslési módszerek megértéséhez ezen felül tulajdonképpen csak a lineáris regresszió alapjainak ismeretére van szükség (az instrumentális változók és a szelekciós tortzítás fogalmát külön bevezetjük). A becslési módszerek technikai részleteit a lábjegyzetekben tárgyaljuk, mint pl. a standard hibák becslésének esetenkénti problémáit, vagy a nemlineáris modellezés részleteit, úgymint a bináris függő változós és időtartam (duration) modellekéit. A legfontosabb megállapítások a lábjegyzetek nélkül is érthetőek.

Az identifikációs probléma miatt Δ_i természetesen nem mérhető: a résztvevők esetében Y_{0i} tényellentétes, a nem résztvevők esetében Y_{1i} tényellentétes. Tehát:

$Y_{1i} \mid D_i = 1$ mérhető, de $Y_{0i} \mid D_i = 1$ nem mérhető; és

$Y_{1i} \mid D_i = 0$ nem mérhető, de $Y_{0i} \mid D_i = 0$ mérhető.

Az első következménye ennek az, hogy Δ_i eloszlása önmagában nem identifikált. Más szóval, egyéb, igen erős feltevések nélkül semmi esélyünk arra, hogy pusztán a megfigyelt eredmények alapján rekonstruáljuk a program hatásának az eloszlását. Ennek az az oka, hogy nem tudhatjuk, aki kezelést kapott és a kezelt csoport eredményváltozó eloszlásában mondjuk magasan helyezkedik el, az kezelés hiányában hol lenne az eloszlásában .

A programok költség-haszon elemzéséhez legfontosabb a hatás (Δ) várható értéke. Ellentétben az eloszlással, a hatás várható értéke bizonyos esetekben elvileg megszorító feltevések nélkül is identifikált. Ennek az az oka, hogy a várható érték lineáris operátor: a különbség várható értéke nem más, mint a várható értékek különbsége. Mint azt majd a későbbiekben bemutatjuk, kísérleti szituációkban becsülhető $E(Y_{1i})$ és $E(Y_{0i})$, és így $E(\Delta_i) = E(Y_{1i}) - E(Y_{0i})$ is.

Az irodalomban kétféle várható értéket is definiálnak: a részvétel vagy "kezelés" várható hatását a népesség (vagy a célcsoport) egy véletlenül kiválasztott tagjára (Average Treatment Effect, ATE), vagy csak a résztvevőkre (Average Effect of the Treatment on the Treated, ATE_1). Elvi szempontokból a két várható hatás között van némi a különbség, praktikusan a kettő azonban általában ekvivalenssé tehető a népesség (célcsoport) megfelelő definiálásával. A várható hatások formális definíciója:

$$\begin{aligned} ATE &= E(\Delta_i) = E(Y_{1i} - Y_{0i}) = E(Y_{1i}) - E(Y_{0i}) \\ &= E(Y_{1i} \mid D_i = 1)\Pr(D_i = 1) + E(Y_{1i} \mid D_i = 0)\Pr(D_i = 0) \\ &\quad - E(Y_{0i} \mid D_i = 1)\Pr(D_i = 1) - E(Y_{0i} \mid D_i = 0)\Pr(D_i = 0) \end{aligned}$$

$$ATE_1 = E(\Delta_i \mid D_i = 1) = E(Y_{1i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 1).$$

A két várható hatás ezek alapján a következőképpen függ össze:

$$\begin{aligned} ATE &= E(\Delta_i \mid D_i = 1)\Pr(D_i = 1) + E(\Delta_i \mid D_i = 0)\Pr(D_i = 0). \\ &= ATE_1 \Pr(D_i = 1) + E(\Delta_i \mid D_i = 0)\Pr(D_i = 0). \end{aligned}$$

A vizsgálatok jelentős részében Y bináris változó (foglalkoztatás). Ilyenkor tehát az a kérdés, hogy a program hatására hogyan változik a foglalkoztatási esély (valószínűség). A fenti várható érték specifikáció ugyanúgy helytálló bináris eredményváltozókra, ha azok 0 – 1 értékekkel kódoltak. Ha 1 a

foglalkoztatottság, 0 a nem foglalkoztatottság, akkor a feltételes várható érték nem más, mint a feltételes valószínűség:

$$E(Y_{mi} \mid akármi) = P(Y_{mi} \mid akármi), \quad m = 0 \text{ vagy } 1.$$

A továbbiakban mindent feltételes várható érték jelöléssel írunk fel, ami bináris eredményváltozó esetén feltételes siker-valószínűségként értelmezhető.

Az ATE esetében két, az ATE_1 esetében egy tényellentétes várható érték szerepel. *E tényellentétes állapotok mérése a hatások identifikációs problémájának az alapja.* Ismét hangsúlyoznunk kell azt a végig használt feltevést, hogy a programnak csak a benne résztvevőkre van hatása. A közvetett hatások mérésével a tanulmány végén külön rész foglalkozik; addig csak a közvetlen hatásokra koncentrálunk.

3. A TÉNYELLENTÉTES HATÁSOK MÉRÉSI MÓDSZEREINEK ÁTTEKINTÉSE

A hatásvizsgálatokban igen sokféle módszert alkalmaznak. Négy típust különböztethetünk meg. Ezek a következők:

- tervezett kísérletek
- természetes kísérletek
- párosított ("matching") modellek
- ökonometriai modellek

A különböző módszerek használata eltérő szituációkban lehet optimális. Általánosságban is elmondható azonban, hogy legalább olyan fontos az, hogy mennyi és milyen minőségű információnk van a megfigyelt egyénekről és azt mennyire vesszük figyelembe a mérésnél, mint az, hogy milyen módszert alkalmazunk (Heckman, Lalonde és Smith, 1999). Ez alól kivételt csak az elsőként tárgyalt tervezett kísérleti módszer képez, mivel ott nem csak az adatok utólagos összegyűjtése és elemzése, de magának a programnak a megtervezése is a mérés céljaival összhangban történik. A program hatásának eloszlását és a hatás mögötti mechanizmusok mérését azonban tervezett kísérleteknél is csak részletes információk megléte esetén lehet elvégezni.

3.1 TERVEZETT KÍSÉRLETEK

Tervezett kísérletekben a programban történő részvétel (az, hogy valaki a kezelt csoportba tartozik, nem pedig a kontroll csoportba) véletlen kiválasztás eredménye. Ideális esetben tehát a két csoport összetétele azonos,

így a két állapot eredményváltozójának az eloszlása jól írja le a tényellentétes eloszlásokat is. Ha F jelöli az eloszlásfüggvényt, akkor:

$$F(Y_{1i} \mid D_i = 1) = F(Y_{1i} \mid D_i = 0) = F(Y_{1i}), \text{ és}$$

$$F(Y_{0i} \mid D_i = 1) = F(Y_{0i} \mid D_i = 0) = F(Y_{0i}).$$

Így tehát

$$E(Y_{1i} \mid D_i = 1) = E(Y_{1i} \mid D_i = 0) = E(Y_{1i}), \text{ és}$$

$$E(Y_{0i} \mid D_i = 1) = E(Y_{0i} \mid D_i = 0) = E(Y_{0i}).$$

A két csoportban mért várható értékek különbsége így identifikálja a számunkra érdekes várható érték különbségeket (amelyek esetünkben megegyeznek):

$$E(Y_{1i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0) = E(Y_{1i}) - E(Y_{0i}) = \text{ATE}$$

$$E(Y_{1i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0) = E(Y_{1i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 1)$$

$$= \text{ATE}_1 \ (=\text{ATE}).$$

Attól hogy $F(Y_{1i})$ és $F(Y_{0i})$ identifikált, $F(\Delta_i)$ még nem lesz az, hiszen ellentétben a várható értékkel, ahol $E(Y_{1i}) - E(Y_{0i}) = E(\Delta_i)$, az eloszlásfüggvényre $F(Y_{1i}) - F(Y_{0i}) \neq F(\Delta_i)$. Az, hogy valaki a kezelt csoportba kerülve az eloszlás tetején található (részvétel esetén az átlagosnál jobb eredményt ér el), nem jelenti azt, hogy a kontroll csoportban is az eloszlás tetején lenne (nem részvétel esetén szintén az átlagosnál jobb eredményt érne el). Általános esetben ugyanis, ha a részvétel nem mindenki számára jár pontosan ugyanolyan hatással ("heterogén hatások"), akkor a részvételtől potenciálisan többet kapók ha valóban a kezelt csoportba kerülnek, felfelé mozdulnak el a kimenetel változó eloszlásában ahhoz képest, ahol kontroll csoportba kerülésük esetében lennének. Minthogy azonban vagy a kezelt, vagy a kontroll csoportban figyelhetünk meg csak valakit, ez az elmozdulás nem mérhető. Heterogén hatások esetén is mérhető azonban az átlagos különbség, de csak az (ez a várható érték speciális – lineáris – tulajdonságából fakad). Kivételes esetben, ha a részvétel mindenki számára pontosan ugyanolyan hatással jár ("homogén hatások"), akkor természetesen a fenti probléma nem áll fenn, így a teljes eloszlás identifikált. A gond csak az, hogy a hatások homogenitása nem vizsgálható közvetlenül. Bővebben lásd Heckman, Smith és Clemens (1997).

Akár homogének, akár heterogének a hatások, a véletlenszerű hozzárendelés ("randomizálás") elvileg megoldja a hatás várható értékének identifikációs problémáját. A tervezett kísérletek megvalósíthatósági lehetőségei azonban igen korlátozottak a társadalompolitikai programok gyakorlatában. Az embereken való kísérletezés nyilvánvaló morális és politikai problémái

mellett praktikus akadályok is nehezítik a kísérletek értékelését. A kísérleti csoportból többen lemorzsolódhatnak és kikerülhetnek a vizsgálat látóköréből, a kontroll csoport tagjai pedig eltérő bánásmódban részesülhetnek, mint a program teljes hiányában. Ez utóbbi gyakori oka az, hogy a program adminisztráló személyzete a véletlenszerűen a kontroll csoportba kerülteket kitüntetetten kezelik, a kísérleti kezelés hiányát kompenzálandó (ezt a jelenséget néhol "helyettesítési hatásnak" is nevezik).

Ha kezelni tudja ezeket a nehézségeket, a véletlen hozzárendelés megoldja a hatás identifikálásának a problémáját, ám a hatásmechanizmusok vizsgálatában közvetlenül nem segít. Ennek az az oka, hogy bár a részvétel véletlenszerű és így nem függ az elérhető eredménytől, a részvételre kondícionált változók (a különböző mechanizmusok) már nem véletlenszerűek. A problémát és a megoldási lehetőségeket részletesen a tanulmány ötödik fejezetében vizsgáljuk, elsősorban Ham és Lalonde (1996) nyomán.

A kontrollált kísérletek egy változatában nem az egyes egyéneket, hanem területi egységeket rendelnek véletlenszerűen a kísérleti és kontroll csoporthoz. Ez a megoldás, bár politikailag nem feltétlenül könnyebb megvalósítani, a praktikus problémákat jobban kezeli. A lemorzsolódás a rendesen futó programokhoz hasonlóan alakul, és így az a teljes hatások értékelését éppen a megfelelő módon befolyásolja. Egész területek részvétele vagy nem részvétele esetén természetesen a "helyettesítési hatás" sem érvényesül, legalábbis nem helyi szinten. E kísérletek problémája nyilvánvalóan azonban az, hogy a "kezelt" és a kontroll területi egységek különböző munkaerőpiacok, és emiatt nagymértékben különbözhetnek egy sor olyan dologban, amelyek befolyásolják az eredményeket a programoktól függetlenül is. Várható értékben a véletlen hozzárendelés ezt kiszűri, adott esetben, főleg kevés területi egység összehasonlításakor azonban semmi nem garantálja a várható érték bekövetkezését.[2]

[2] Elvileg a probléma az egyéni kísérleteknél is felmerül: ott is egy sor egyéni dolog befolyásolja a kimemetelt. Sok száz vagy sok ezer kísérleti résztvevő esetében azonban jobban kiátlagolódnak ezek az egyedi jellemzők, mint néhány területi egységgel történő kísérlet esetén. Statisztikailag megfogalmazva, bár ideális esetben mindkét megközelítés tortzítatlanul becsüli az átlagot, a területi egységek véletlen kiválasztásával becsült átlagnak jóval nagyobb a varianciája, mert az jóval kevesebb megfigyelés között randomizál.

3.2 TERMÉSZETES KÍSÉRLETEK

A természetes kísérleteken alapuló elemzések tipikusan a területi egységek véletlenszerű hozzárendelését veszik mintául. Esetükben a programban való részvétel nem véletlenszerű, ám az elemzők azzal a feltevéssel élnek, hogy az, hogy hol indult ilyen program és hol nem, az független attól, hogy a milyen potenciális eredménnyel jártak volna. Tehát a kísérlet ugyan nem kontrollált, de megbízunk a természetben – esetünkben a társadalompolitikai döntéshozókban és a döntések végrehajtóiban –, hogy a potenciális eredményektől függetlenül jártak el.[3]

Mint azt már fentebb megjegyeztük, a különböző területen élőkből álló kezelt és kontroll csoportok összehasonlításának, így természetes kísérleteknek egy nagy hátránya az, hogy az összehasonlított egyének nem azonos munkaerőpiacon vannak. Ez felveti azt a kérdést, hogy az eredmények különbségei mennyiben tudhatók be a program hatásának, és mennyiben az eltérő munkaerőpiaci változásoknak. Általános esetben a különböző területi egységek kezdőállapotbeli eredményváltozói (pl. átlagos vagy csoportonkénti munkanélküliségi ráta) nem azonosak, így természetesen nem a program utáni eredményeket, hanem azoknak a kezdőállapottól való különbségeit kell összehasonlítani. Ez külön nevet is kapott: "difference-in-differences" (DID). A DID megközelítés tehát nem az eredményeknek, hanem azok változásának a függetlenségét követeli meg hozzárendeléstől. Az irodalomban a DID elnevezést általában a természetes kísérletek eredményeit összehasonlító modellekre használják, a kontrolláltakéra nem. Természetesen a kezdőállapotra való kontrollálás önmagában nem oldja meg e különböző munakerőpiacok összehasonlíthatóságának a problémáját, hiszen lehet, hogy a változások is különböztek volna a program teljes hiányában is. A legegyszerűbb DID stratégia alapvető identifikációs feltevése az, hogy a változások ugyanakkorák lettek volna.

Elvileg kevésbé restriktívek azok a természetes kísérleti modellek, ahol az eredmények függetlenségét csak megfigyelhető változókra kondícionálva követelik meg. Ezek felfoghatók az alább tárgyalt párosítási és

[3] Ez a feltevés magukból az adatokból nem, csak esetleg külső információk alapján ellenőrizhető. Centralizált döntések esetén a feltevésnek az a tartalma, hogy a döntéshozók vagy nem képesek felmérni a potenciális hatásokat (még közelítőleg sem), vagy ez egyáltalán nem érdekli őket a döntés meghozatalánál. Normálisan működő országok társadalompolitikai gyakorlatában azért ez igen erős feltevés. Nem centralizált döntéshozás esetén természetesen mindez kevésbé restriktív. Ezért is gyakoribb természetes kísérleti sziutációnak minősíteni területi különbségeket federális berendezkedésű országokban (USA), vagy esetleg egyébként igen hasonló országok között.

ökonometriai modellek speciális eseteinek, ezért részletesen itt nem foglalkozunk velük.

3.3 PÁROSÍTÁSI (MATCHING) MODELLEK

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 reprodukálja Δ_i-t, páronként. Természetesen sosem lehet tökéletesen reprodukálni Δ_i-t, hiszen az ugyanazon egyén két állapotbeli eredményének a különbsége. Elvileg azonban közel tudunk kerülni ehhez a célhoz akkor, ha a párok tagjai minden olyan változó tekintetében hasonlóak, amelyek befolyásolják az eredményt akármelyik állapotban.

Jelöljük az i-ik egyén megfigyelhető ismérveit X_i vektorral (a vektorokat általában nem jelöljük eltérően a skalároktól, csak akkor, ha ez fontos a megértéshez). Formálisan a matching modellek legfontosabb feltevése az, hogy X-re kondícionálva a kontroll állapot eredménye független a részvételtől:

$$Y_{0i} \perp D_i \mid X_i$$

Ebben az esetben a résztvevők tényellentétes eredményeit (amit részvétel nélkül értek volna el) jól reprodukálják azoknak a nem résztevőknek az eredményei, akik ugyanazokkal a megfigyelhető ismérvekkel rendelkeznek. Ez a feltevés nyilvánvalóan nagyon erős. Annál inkább hihető, minél több mindent tudunk megfigyelni az emberekről (mind a résztvevőkről, mind a nem résztvevőkről: X a mindkét csoportban megfigyelhető változók vektora).

A másik fontos feltevés az, hogy minden résztvevőnek van nem résztvevő megfelelője. Ez a feltevés még ha a populációban igaz is, igen ritkán teljesül véges mintákban, különösen ha a változók száma nagy (ami viszont általában szükséges az előző feltevés legalább közelítőleges teljesítéséhez is). A legsúlyosabb esetben populációban sem lehet megfelelő párokat találni a résztvevőkhöz. Ilyenek az általános jellegű programok: ezekben mindenki automatikusan részt vesz a programban ha bizonyos kritériumoknak eleget tesz (pl. jövedelem vagy etnikai hovatartozás).

Ha elvileg találhatók megfigyelhető tulajdonságaikban azonos (vagy nagyon hasonló) kontroll párok, véges mintában akkor sem feltétlenül sikerül minden résztvevőnek párt találni. Ez azért probléma, mert információt vesztünk ha a minta nem minden egyedét használjuk a méréshez. A probléma legelterjedtebb megoldása az úgynevezett “propensity score” (hajlandósági mutató) alapú párosítás. Ezt a módszert Rosenbaum és Rubinstein

(1983, 1984) vezette be. A propensity score nem más, mint a programban való részvétel megfigyelhető ismérvek alapján történt becsült valószínűsége:

$$p(X_i) = P(D_i = 1 \mid X_i)$$

Rosenbaum és Rubin (1983) kimutatták, hogy ha X-re kondicionálva a kontroll állapot eredménye független a részvételtől (ami a párosítási modellek alapvető feltevése), ez implikálja azt, hogy a kontroll állapot eredménye a propensity score-ra kondicionálva is független a részvételtől:

$$Y_{0i} \perp D_i \mid X_i \Rightarrow Y_{0i} \perp D_i \mid p(X_i).$$

Ezt az állítást propensity score tételnek is nevezik. Másik fontos eredményük az, hogy a részvétel esemény kondicionálva az propensity score-ra független X-től:

$$X_i \perp D_i \mid p(X_i)$$

Ennek az a következménye, hogy az azonos propensity score-ral rendelkező megfigyelések esetén a megfigyelhető változók (X) eloszlása ugyanaz a kezelt és a kontroll csoportban. Tehát azonos propensity score-ral rendelkező kezelt és kontroll egyedeknek elég egyszerűen összehasonlítani az eredményüket, az X további kontrollálására nincs szükség. A párosításhoz magához természetesen szükség van arra, hogy az adott megfigyelhető változókkal (X) rendelkezők között legyen a kezelt és a kontroll csoportban is egyén. Sokaságban megfogalmazva ez azt jelenti, hogy azok az X értékek, amelyek valakit egyértelműen részvételre vagy nem részvételre determinálnak, nem szolgáltathatnak információt a kezelés hatásának becsléséhez. Technikailag ez a feltevés ekvivalens azzal, hogy a propensity scorenak a nulla-egy intervallum belsejében kell lennie:

$$0 < p(X_i) < 1.$$

A propensity score használatával egyetlen folytonos változó alapján kell a párosítást elvégezni. Ezt kétféle módon szokták megtenni: vagy minden részvevőhöz egyetlen párt rendelnek, amely a legközelebb van hozzá, vagy egy előre megadott távolságon belül mindenkit hozzárendelnek (de potenciálisan eltérő, például a távolsággal csökkenő súllyal).[4]

[4] Technikailag az ilyen párosítási becslések két lépésből állnak. Először valamilyen bináris függő változós modellel (tipikusan flexibilis, tehát a változók magasabbrendű polinómjait és interakcióit is tartalmazó logit vagy probit) megbecsülik a részvétel valószínűségét. Az így kapott becsült valószínűség, az empirikus propensity score eloszlását ezután intervallumokra bontják, és ezen belül végzik el a párosítást (egy az egyhez, egy a többhöz, vagy több a többhöz). Az eredményeket ezekben a párokban (több-a-többhöz párosítás esetén lényegében alminta-párokban) összehasonlítják ösz-sze. Az átlagos hatást (ATE, ATE_1) ezeknek az átlagolásával (nem egy-az-egyhez párosítás esetén súlyozott átlagolásával) kaphatjuk meg.

A párosítási modellek hátránya, hogy a kétlépcsős nemstandard becslési eljárás miatt a hatások becslésének mintavételi hibáját nehéz megbecsülni.[5] Nagy előnye viszont az, hogy ha elhisszük a párosítás érvényességéhez szükséges feltevéseket, egyéb feltevések nélkül (nemparaméteresen, vagy a propensity score használata esetén félparaméteresen) becsülnek. Vagyis nem kényszerítenek rá önkényes függvényformákat a modellekre. Ennek abban az esetben van jelentősége, ha a hatások nem ugyanazok mindenkinél (minden megfigyelhető *X* esetében). Ilyenkor ugyanis a párosítás segítségével nemcsak az átlagos hatás, hanem annak eloszlása is becsülhető. Másik nagy előnye, hogy automatikusan csak összehasonlítható kezelt és kontroll egyéneket használ a becsléshez – ennek jelentőségét a következő fejezetben tárgyalt ökonometriai modellekkel összehasonlítva érthetjük meg.[6]

3.4 ÖKONOMETRIAI MODELLEK

Az ökonometriai modellek közül két nagy csoportot tekintünk át: a keresztmetszeti regressziókat és szelekciós modelleket. Ezekbe szinte valamennyi olyan modell belefér, amit a munkanélküli programok hatáselemzéséhez használni szoktak. Más gazdaságpolitikai döntéseknek (pl. adószabályok változásainak) a társadalmi hatásához gyakran építenek fel bonyo-

---

[5] A standard hibára aszimptotikus közelítést kaphatunk a delta-módszerrel. Kimutatható, hogy ha a részvételi egyenlet (a becslési eljárás első lépése) konzisztensen becsüli a részvétel valószínűségét, akkor az ATE következő becslőfüggvénye a hatásos (minimális aszimptotikus varianciájú): $N^{-1}\Sigma_i Y_i[D_i - p^*(X_i)]/\{p^*(X_i)[1- p^*(X_i)]\}$ – ahol N az összesített minta elemszáma, az i index az összesített mintában az egyes megfigyeléseket (egyének) jelöli, és $p^*(X_i)$ az i-ik egyénre becsült propensity score. Az eredményt Hirano, Imbens, és Ridder (2000) bizonyította Hahn (1998) alapján. A részvétel konzisztens becsléséhez általános esetben, tehát a feltételes valószínűségi függvény a priori specifikációja nélkül, nemparametrikus becslésre van szükség. A sok magyarázó változós nemparametrikus bináris függő változós modellek becslése azonban rendkívül nehézkes, egy bonyolultsági fok után pedig praktikusan kivitelezhetetlen. A szakma jelenlegi álláspontja szerint valószínűleg nem veszítünk sokat a hatásosságból, ha az első lépcsőben rugalmas parametrikus modellt (pl. logit magasabbrendű polinómokkal és interakciókkal) használunk (Wooldridge, 2002, 18.3. fejezet).

[6] Egy utolsó megjegyzés a propensity score alapú párosítási modellekhez. A módszer kidolgozói nem közgazdászok voltak, és eredményeiket nem az ökonometriai szakmán belül publikálták (Rosenbaum és Rubin, 1983, például a vezető biometriai folyóiratban jelent meg). A módszer közgazdaságtani alkalmazása csak az 1990-es évek közepén kezdett elterjedni, éppen a társadalompolitikai programok hatásvizsgálatában. Napjainkra lett igazán divatos az ökonometria és alkalmazott közgazdaságtan művelői körében, de a programok hatásvizsgálatának területén kívül továbbra sem nagyon alkalmazzák.

lult strukturális modelleket, ám ezek egyrészt specifikusak az adott kérdésre, másrészt kevéssé használatosak a mi témánkban.

3.4.1 Keresztmetszeti regressziók

A legegyszerűbb ökonometriai modellek az egyéni szintű keresztmetszeti regressziók. Ezek hasonló feltevésen alapulnak mint a párosítási modellek: a megfigyelhető ismérvekre (X) kondicionálva a részvétel (D) független a lehetséges eredményektől (vagy legalábbis korrelálatlan velük). Ezek között is a legegyszerűbb modellben csak az átlagos hatás becsülhető:

$$Y_i = \beta' X_i + \alpha D_i + U_i$$

ahol α becsüli az átlagos hatást (ATE).[7] U a nem megfigyelt változók azon részét foglalja magában, amelyek hatással vannak Y-ra ("nem megfigyelt heterogenitás" Y változóban). Az átlagos hatás (α) identifikálásának legfontosabb szükséges feltétele a részvétel egzogenitása (a függetlenségnél kicsit enyhébb feltétel):[8,9]

$$\mathrm{E}(U_i) = 0,$$

$$\mathrm{E}(D_i U_i) = 0.$$

A modell linearitásában és abban különbözik a párosítási modellektől, hogy csak az átlagos hatást becsüli. A linearitás természetesen nem annyira restriktív elvileg, hiszen dummyk, polinómok, szakaszos paraméterek ("linear splines"), és ezek interakciói is szerepeltethetők az egyenletben. Amennyi-

---

[7] Bináris eredményváltozó esetén is becsülhető ilyen lineáris regresszió, aminek külön neve is van: lineáris valószínűségi modell (Linear Probability Model, LPM). Az LPM-nek előnyei mellett – egyszerű a becslése és a becsült paraméterek interpretációja – sok hátránya van (a prediktált valószínűségek nem korlátozódnak a [0,1] intervallumra, a hibatagok heteroszkedasztikusak és GLS becslés pont az előző probléma miatt nem mindig lehetséges, stb.). Tipikusan ezért ilyenkor – főleg alacsony vagy magas, tehát nem 0.5 körüli valószínűségű események esetén – nemlineáris, úgynevezett index modelleket (logit, probit) használnak.

[8] Szükséges még minden olyan X egzogenitása is, amely korrelált D-vel. Ez praktikusan csak az erősen korrelált X-ekre kell hogy teljesüljön, hiszen gyenge korreláció esetén a torzítás elenyésző. A nagyon erősen korrelált X-ek esetében (pl. szinte csak nők vesznek részt a programban) ez probléma, ha ezek az X-ek nem egzogének, vagyis összefüghetnek a program sikerét mérő változó nem megfigyelhető komponenseivel (pl. azzal, hogy mennyire szeretne az egyén karriert építeni, hiszen a sok gyermeket tervező nők általában kevésbé terveznek hosszú és csak a munkának éléssel járó karriert). Ilyenkor járható út a minta restrikciója is (a példánkban csak nőkre), hiszen ekkor a változó már nem szerepel a modellben.

[9] Nemlineáris modelleknél a korrelálatlanságnál némileg erősebb az egzogenitás feltevés: Cov(U,X)=0 helyett E[U|X]=0 (átlag-függetlenség). Ez U várható értékének X-től való nemlineáris függését is kizárja.

ben azonban sok változónk van (és ez általában szükséges ahhoz hogy hihető legyen a részvétel egzogenitása), akkor ezek a rugalmas függvényformák nem alkalmazhatóak korlátok nélkül a fellépő multikollinearitások és főként a becsülendő paraméterek nagy száma miatt.

Az X-ek között lehetnek a program indulása (vagy résztvevőknél a részvétel kezdete) előtti adatok is. Ekkor technikailag keresztmetszeti modellről van ugyan szó, az időbeliség azonban tartalmilag megjelenik a regressziókban. Az eredményváltozó maga is lehet változás. Például, ha a jövedelmekre gyakorolt hatás a kérdés, akkor Y jelentheti a jövedelem változását, így egy előtte-utána összehasonlításból identifikálható a program hatása. Ez utóbbi gyakorlatilag egy DID modell, mindenféle megfigyelhető változóra (X) is kontrollálva.

Kicsit bonyolultabb specifikációval elvileg mérhető az is, hogy a hatás hogyan függ össze a különböző megfigyelhető változókkal (csakúgy mint a párosítási modellekben, csak megint a lineáris restrikciókkal):

$$Y_i = \beta' X_i + \alpha D_i + \gamma'(D_i X_i) + U_i .$$

A részvétel interakciójával elvileg rekonstruálhatók az X-enként változó hatások, azonban praktikusan megint könnyen belefuthatunk a túl sok becsülendő paraméter problémájába. A lineáris regresszió tehát, bár elvileg tetszőlegesen rugalmassá tehető, a gyakorlatban valóban restriktívebb a párosító modelleknél. Jóval egyszerűbb azonban a becslése. Az előnyök és hátrányok csak a konkrét kutatási kérdések ismeretében mérlegelhetők.

A párosítási modellektől eltérően a regressziók nem követelik meg, hogy a résztvevők és a többiek esetében ugyanaz legyen a megfigyelhető változók (X) értelmezési tartománya. A regressziók becsülhetők akkor is, ha van olyan X tartomány, amelyben csak résztvevők (vagy csak nem résztvevők) találhatók. Ha a (paramétereiben) lineáris specifikáció helyes, akkor ez nem torzítja a hatás (α) becslését, csak az eredmények extrapolálhatóságát teszi kérdésessé. A lineáris specifikáció azonban általában csak közelítés, nem pedig a tudottan jó függvényforma. Ekkor a nem teljesen közös értelmezési tartományú X-ek esetében más β-kat identifikálhatunk mint ha a modellt csak a közös X-ekkel rendelkezőkre futtatnánk le, és ez a hatás paramétert (α) is torzíthatja. Ezért, bár első látásra nem szükséges, érdemes a regressziós modelleket is csak olyan nem résztvevők adatain futtatni, akik a résztvevőkhöz hasonlóak megfigyelhető ismérveikben. Így egyébként a parametrikus függvényforma restriktív (ám egyszerűsítő) feltevésén túl igen jól reprodukálhatók a párosítási modellek eredményei (legalábbis az átlagos hatásra vagy annak egyes X-ekkel való változására vonatkozóan).

3.4.2 Szelekciós modellek

Az ökonometriai modellek bonyolultabb (és ambíciózusabb) változatai nem feltételezek a részvétel egzogenitását, hanem magát a részvételt is modellezik. Ezek a modellek felfoghatók úgy, mint az egyszerű keresztmetszeti regressziók kiegészítése a részvétel valamilyen modelljével. A szelekciós modellek a részvétel egzogenitását leszámítva tehát rendelkeznek a regressziók mindazon tulajdonságával, amelyeket fent tárgyaltunk.

A legegyszerűbb szelekciós modell a kétegyenletes instrumentális változós (IV) modell:

$$Y_i = \beta' X_i + \alpha D_i + U_i$$

$$D_i = \gamma_1' X_i + \gamma_2' Z_i + V_i$$

$$\mathrm{E}(U_i) = 0,\ \mathrm{E}(U_i|X_i) = 0,\ \mathrm{E}(U_i|Z_i) = 0,$$

$$\mathrm{E}(V_i) = 0,\ \mathrm{E}(V_i|X_i) = 0,\ \mathrm{E}(V_i|Z_i) = 0,$$

$$\text{de } \mathrm{E}(V_i|U_i) \neq 0 \text{ és ezért } \mathrm{E}(U_i|D_i) \neq 0.$$

Z olyan változó (vagy változók egy vektora), amely nem része X-nek. Az instrumentumok[10] (Z) tehát olyan változók, amelyek összefüggnek a részvétellel, ám azon kívül nem befolyásolják az eredményváltozót.[11]

Az IV modellek lényege az, hogy a részvétel hatását csak ott vesszük bele a mérésbe, ahol a részvételt az instrumentum okozta. Technikailag nem mindig pontos, ám a megértést segíti az a megfogalmazás, hogy az IV modellek leszűkítik a mintát azokra a résztvevőkre és nem résztvevőkre, akik az instrumentum hatására kerültek egyik vagy másik csoportba. A részvételi döntést megváltoztató instrumentum (Z) akkor érvényes, ha ezen túl más, közvetlen hatása nincs Y-ra.

Amennyiben a program hatása mindenkire ugyanakkora ("homogén"), akkor az IV identifikálja azt (és így természetesen az átlagos hatást – ATE – is). Ha viszont nem ugyanaz a hatás mindenkire ("heterogén hatások"), akkor egyáltalán nem biztosított, hogy az IV identifikálja az átlagos hatást.

---

[10] Tágabb értelemben minden egzogén változó instrumentum, így az X-ek is. A továbbiakban azonban csak Z-t (Z-ket) hívjuk instrumentum(ok)nak, tehát az összes egzogén változó közül az(oka)t, amely(ek) nem függ(nek) össze a program eredményével (Y).

[11] Meg kell jegyeznünk, hogy a fenti felírásban D egyenlete (a második egyenlet) egy lineáris valószínűségi modell. Ez a legegyszerűbb olyan modell, ami bináris (0 vagy 1) kimenetelt magyaráz, ám az ökonometriai alkalmazásokban hagyományosan nem preferálják (helyette inkább index modelleket, pl. probitot használnak). Ehhez az alkalmazáshoz a lineáris valószínűség általában mégis megfelelő, de természetesen lehetséges a második egyenletet más bináris függő változós modellel is felírni. Ez a becslési eljárást bonyolítja, de az intuitív tartalmon nem változtat.

Ennek oka az identifikáció leszűkítése azokra, akik Z miatt vesznek vagy nem vesznek részt a programban. Általában ugyanis semmi sem indokolja, hogy a Z miatt másképp résztvevőkre a programnak ugyanaz legyen a hatása (akár átlagosan is), mint azokra, akiket Z nem befolyásol. Heterogén hatások esetén ezért az IV csak lokálisan identifikálja az átlagos hatást: az ATE helyett az úgynevezett LATE-t (local average treatment effect, Imbens és Angrist, 1994 nyomán). A LATE interpretáció nagy problémája az, hogy általában nem tudjuk pontosan meghatározni a lokációt, vagyis hogy kik azok akiknek a részvételére hat(na), és kik azok, akik részvételére nem hathat(na) az instrumentum. (Angrist, Imbens és Rubin, 1996, és Heckman, 1996).

A fenti kétegyenletes IV modell feltételezi, hogy a megfigyelhető változók (X) ugyanúgy hatnak az eredményváltozóra (Y) a résztvevők esetében, mint a nem résztvevőknél. Ezt a restrikciót oldják fel az úgynevezett "switching" regressziós modellek (Willis és Rosen, 1979), amelyekben a két állapotra (részvétel és nem részvétel) külön specifikált egyenlet írja le a megfigyelhető ismérvek és az eredményváltozó kapcsolatát.

$$Y_{1i} = \beta_1 \,'X_{1i} + U_{1i}$$

$$Y_{0i} = \beta_0 \,'X_{0i} + U_{0i}$$

$$D_i = \gamma_1 \,'X_i + \gamma_2 \,' Z_i + V_i$$

Az eddigi jelöléseknek megfelelően Y_1 a résztvevők eredményét, Y_0 pedig a nem résztvevőkét jelenti. A modell megengedi (de nem követeli meg), hogy a két állapotban különböző megfigyelhető változóknak különböző legyen a hatása (nem csak β_1 és β_0 lehet eltérő, de az X_1 és X_0 vektorokban is lehetnek különböző változók). A részvételi egyenletben szereplő X tartalmazza X_1 és X_0 elemeit, de nem szükségképpen mindet, míg Z olyan változó (vagy vektor), amely közvetlenül nem befolyásolja az eredményeket, csak a részvételt. Z tehát megfelel a kétegyenletes IV modell instrumentumának. A modell felírásában közvetlenül nincs ott az átlagos hatás paramétere: ez a két eredmény-egyenlet konstansainak a különbsége (vagyis $\beta_1 - \beta_0$ egyik eleme).

A harmadik egyenletet hagyományosan nem lineáris valószínűséggel, hanem index modellel szokás felírni:

$$\mathrm{P}(D_i = 1) = \mathrm{F}\,(\gamma_1 \,'X_i + \gamma_2 \,' Z_i + V_i\,),$$

ahol F valamilyen eloszlásfüggvény (standard normál ha probit, logisztikus ha logit).

A modell feltevései, analóg módon a kétegyenletes IV modellhez:

$E(U_{1i}) = 0$, $E(U_{1i}|X_{1i}) = 0$, $E(U_{1i}|Z_i) = 0$,

$E(U_{0i}) = 0$, $E(U_{0i}|X_{0i}) = 0$, $E(U_{0i}|Z_i) = 0$,

$E(V_i) = 0$, $E(V_i|X_i) = 0$, $E(V_i|Z_i) = 0$,

de $E(U_{0i}|V_i) \neq 0$ és/vagy $E(U_{1i}|V_i) \neq 0$ és ezért $E(U_{0i}|D_i) \neq 0$ és/vagy $E(U_{1i}|D_i) \neq 0$.

Becslési részleteit tekintve ez a szelekciós modell sokban különbözik az egyszerű kétegyenletes IV modelltől, intuitív tartalma azonban azonos (leszámítva a kevesebb restrikciót az X-eknek Y-ra gyakorolt hatását illetően). Ezért érvényes rá a mindaz, amit az IV modellekről leírtunk. Felelevenítve azokat, a modell legfontosabb tulajdonságai a következők. A kezelés egzogenitását nem követeljük meg, ám helyette az instrumentum(ok) egzogenitása a feltétel.[12] Általában igen nehéz ilyen instrumentumokat találni. Azután, ha a hatások nem homogének, akkor a switching modell is csak lokálisan identifikálja az átlagos hatást (LATE). Végül, minthogy annak kiterjesztéséről van szó, a regressziós modellek problémáinak jó része is érvényes marad. Ezek közül talán legfontosabbak a függvényforma restrikciók és a két csoport megfigyelhető változóinak összehasonlíthatóságának kérdései.

3.5 Érzékenység-vizsgálat és robusztus határok

A becslési modellek mindegyike valamilyen feltevés-rendszeren alapul. Tervezett kísérleteknél a helyettesítési hatás hiánya, a többi módszernél a hozzárendelés (vagy szelekciós modellekben az azt meghatározó instrumentum) függetlensége a legfontosabb feltevés. Igazán körültekintő hatásvizsgálatnak ki kell térnie nemcsak arra, hogy miért gondoljuk úgy, hogy az alkalmazott modell mögötti feltevés a leginkább valószínű, hanem arra is, hogy hogyan változnak a legfontosabb eredmények, ha a feltevések (vagy azok egy része) nem, vagy nem teljes mértékben állnak. Az ilyen elemzéseket nevezik érzékenység-vizsgálatnak ("sensitivity-analysis"). Az érzékenység-vizsgálat egy példája, amikor egy egyszerű propensity score becslésen (pl. a magyarázó változóiban lineáris logit) alapuló eredményeket bonyolultabb módszerrel (magasabbrendű polinómok, interakciók a

[12] A switching modellek hagyományos formájában az eredmény-egyenletek lineárisak, a részvételi azonban nem (probit). Ezért techninkailag identifikált a modell akkor is, ha nincsen benne egyetlen instrumentum (egzogén Z) sem. Minthogy ez az identifikáció pusztán a függvényformákon alapul, általában nehezen hihető.

logitban) újrabecsülik, és megvizsgálják, mennyire változnak a legfontosabb eredmények.

Az érzékenység vizsgálatok mellett az eredményeknek az identifikáló feltevéstől való függését elvileg robusztus határok meghatározásával is lehet vizsgálni. A megközelítés – amely általánosságban minden identifikációs probléma esetén alkalmazható – lényege, hogy a hatások megbecsülhetők extrém feltételezések mellett. Nem nyilvánvalóan véletlenszerű szelekció esetén például becsülhetők úgy is, hogy feltesszük, a szelekció véletlenszerű volt, és úgy is, hogy feltesszük, csak azok vettek részt a programban, akiknek a legtöbb hasznot hozta. Az extrém feltételek behatárolják a hatás lehetséges intervallumát. Az elvi vonzereje ellenére a módszert viszonylag ritkán alkalmazzák, mivel tipikus esetben az így becsült intervallumok túl tágak és ezért nem informatívak.

3.6 A MÉRÉSI MÓDSZEREK RÖVID ÖSSZEHASONLÍTÁSA

A bemutatott módszerek mindegyikének vannak előnyei és hátrányai. A kontrollált kísérletek a véletlenszerű hozzárendelés miatt megnyugtató megoldást nyújtanak az identifikációs problémára, ám morális és praktikus akadályai vannak. Ráadásul a hatásmechanizmusok megértésében felmerülő identifikációs problémákat a véletlen hozzárendelés önmagában nem oldja meg (lásd a következő fejezetet). A természetes kísérleteknél nagyon hasonló a helyzet, két kivétellel. A részvétel véletlenszerűsége nem biztosított automatikusan, hanem az a kutató feltételezése. E feltevés hihetősége sokszor megkérdőjelezhető, ráadásul a modellen belül nem tesztelhető maradéktalanul. Amennyiben azonban elfogadjuk, a természetes kísérleteknek nem lebecsülendő előnye, hogy a kísérletezés morális és politikai problémái nem merülnek fel: a kísérlet a kutatástól függetlenül lezajlott. A párosításon alapuló és az ökonometriai modelleknél szintén nem biztosított a részvétel függetlensége a potenciális eredményektől, hanem az a kutató feltételezéseinek a tárgya. A párosítási modellek előnye, hogy nem kényszerítenek rá semmilyen előre meghatározott formájú függvényt a megfigyelhető ismérvek és az eredmények kapcsolatára. Ezen kívül a hatásoknak nem csak a várható értékét, hanem az eloszlását is mérhetővé teszik. Hátrányuk azonban az, hogy különösen nagy mintákat igényelnek, és a párosítási eljárások gyakran igen bonyolulttá válnak. Az ökonometriai modellek az egyszerűbb becsülhetőséget restriktívebb függvényformák alkalmazásával érik el, de ezáltal elrejtenek olyan problémákat is, amelyek a párosítási modelleknél explicit módon megjelennek. Amennyiben azonban tisztában vagyunk ezekkel és megfelelően odafigyelünk rájuk, ez utóbbiak általában kezelhetők az ökonometriai modellek esetében is.

Mindegyik módszerre érvényes, hogy a rendelkezésre álló adatok mennyisége és minősége kritikus módon befolyásolja a hatásvizsgálat megbízhatóságát. A párosítási és ökonometriai modelleknél ez nyilvánvaló. Ezeknél ugyanis vagy arra van szükség, hogy a kutató minden olyan információval rendelkezzen ami befolyásolja a részvételt és az eredményeket, vagy megfigyeljen egy olyan mechanizmust, amiről tudja (feltételezi), hogy befolyásolja a részvételt, de az eredményeket közvetlenül nem (ez az instrumentum). A kontrollált és természetes kísérletek hatásának az elemzéséhez elvileg jóval kevesebb információ elegendő, több okból azonban itt is fontos a megfelelő mennyiségű és minőségű adat. Egyrészt szükséges ez a kísérletek véletlenszerűségének és az esetleges mellékhatásoknak a vizsgálatához. Másrészt, megfigyelhető változókra kondícionálva a hatások átlaga mellett mérhetővé tehető az azokkal összefüggő eloszlás is. Harmadrészt, a hatásmechanizmusok feltárásánál az eredetihez hasonló identifikációs problémákba ütközünk, amelyeket már nem old meg a részvétel véletlenszerűsége.

A szükséges mérhető ismérvek és nem kísérleti szituációkban a kontroll csoport megválasztásának problémáját tárgyalja a következő rész. A hatásmechanizmusok feltárásánál felmerülő nehézségeket és lehetséges leküzdésüket pedig az azt követő rész mutatja be.

4. A KONTROLL CSOPORT MEGVÁLASZTÁSÁNAK KÉRDÉSEI NEM KÍSÉRLETI SZITUÁCIÓBAN

Amennyiben a programban való részvétel nem véletlenszerű kiválasztás eredménye (vagy, a "természetes kísérletektől" eltérő módon, nem tételezhető fel a kimeneteltől független részvétel), a résztvevők tényellentétes eredményeit a kutató által kiválasztott csoport eredményeiből kell identifikálni. Elvileg igen egyszerű eldönteni, hogy milyen a megfelelő kontroll csoport: tagjai megfelelnek minden kritériumnak ami a programban való részvételhez szükséges, ugyanazon vagy legalábbis hasonló élethelyzetben és munkaerőpiacon vannak (vagy lehetnek), de mégsem vesznek részt a programban. Praktikusan azonban ilyen kontroll csoportok kiválasztása nagy nehézségekbe ütközhet.

Ebben a fejezetben ezekkel a kérdésekkel foglalkozunk. Először azt vizsgáljuk, általában miért nem szerencsés a részvételi kritériumoknak nem megfelelő egyénekből álló kontroll csoportot használni a tényellentétes eredmények becsléséhez. Ezután LaLonde (1986) nagy hatású tanulmányán keresztül azt vizsgáljuk, hogy milyen következményekkel jár, ha országosan reprezentatív mintákat használunk kontroll csoportnak. Érvelünk

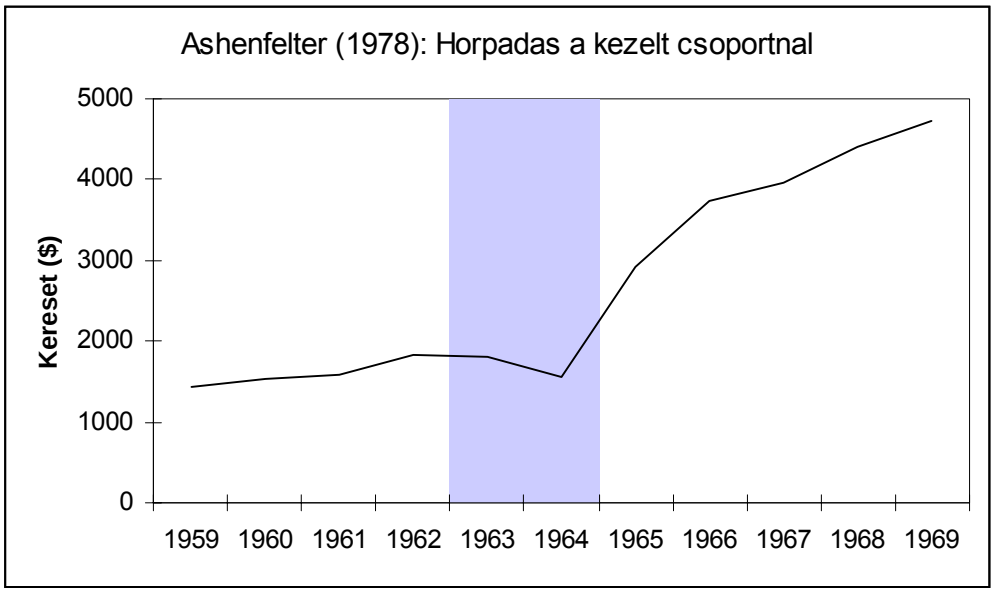
amellett, hogy a becslési módszer és a kontroll csoport kiválasztása nem két egymástól elkülönülő kérdés, hanem egyazon tartalmi problémának a két oldala. Ezt támasztja alá az utána következő alfejezet, amelyben a párosítási modellek előnyeit vizsgáljuk azonos szituációban. Ebben kiderül, hogy nem feltétlenül az különbözteti meg a nem kísérleti eredményeket, hogy párosítási vagy ökonometriai modelleket használunk-e, hanem az, hogy azok milyen kontroll csoportot definiálnak az összehasonlításhoz. A fejezet végén összegezzük mindezek implikációit.

4.1 Kontrollok és a program részvételi kritériumai

Az első nyilvánvaló probléma abból adódhat, ha az adott gazdaságban mindenki részt vehet a programban, ha bizonyos kritériumoknak megfelel. Az ezeknek a kritériumoknak nem megfelelő egyének nyilván nem jó kontrollok, de azok sem tökéletesek, akik megfelelnek nekik, mégsem vesznek részt. A részvétel ugyanis döntés kérdése.

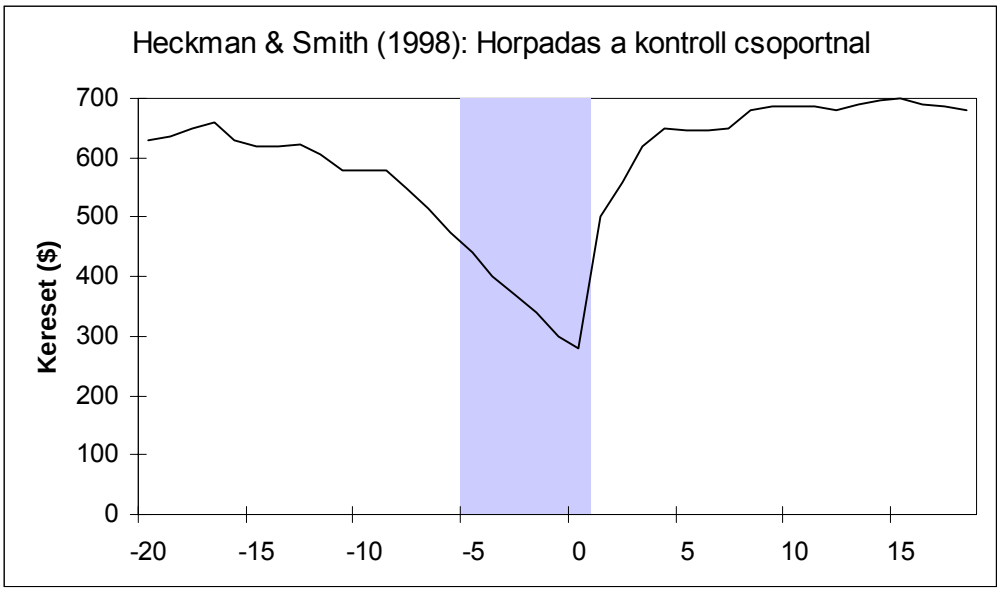
Tekintsünk egy igen elhíresült okot arra, hogy miért nem feltétlenül szerencsés a kritériumoknak nem pontosan megfelelő egyéneket kontroll csoportként használni. A jelenséget első felfedezőjéről, Ashenfelter (1978)-ról "Ashenfelter-horpadásnak" (Ashenfelter's dip) nevezik. Lényege az, hogy az aktív foglalkoztatási programokban való részvételnek általában feltétele valamilyen hátrányos munkaerőpiaci helyzet. Tipikusan az egyénnek munkanélkülinek vagy inaktívnak kell lennie az indulás időpontjában és azt néhány hónappal megelőzően. A programban részt vevők ezért szelektált mintája lesz a nagy valószínűséggel hátrányos helyzetbe kerülők (állásukat nagy valószínűséggel elvesztők) sokaságának is: ők azok, akikkel pont a megfelelő időszakban történt meg a negatív esemény. A kritériumként szolgáló időszak előtt azonban egy részük nem volt hátrányos helyzetben. Ugyanilyen okok miatt a program kritériumainak megfelelő egyének közül néhányan ki tudnának lábolni ebből a helyzetből a program nélkül is. A következő két ábra két példát mutat az Ashenfelter-horpadásra: az első Ashenfeltertől (1978) való, és egy hagyományos átképzési programban résztvevők éves keresetét mutatja (mindenféle inflációs igazítás nélkül). A képzés 1964-ben kezdődött, és akkor kellett megfelelni a munkanélküliség követelményének. A képzés lezárulta után (1965-től kezdve) megfigyelhetjük a (nominális) keresetek jelentős emelkedését. A probléma az, hogy csak a résztvevők kereseteit vizsgálva nem tudhatjuk, ebből az emelkedésből mennyi a program hatása, és mennyi az autonóm növekedés.

1. ábra: **Ashenfelter-horpadás egy hagyományos átképzési program résztvevői esetén** (Ashenfelter, 1978)



Autonóm növekedésre példa a *2. ábra*, amely Heckman és Smith (1998) alapján egy kísérleti kontroll csoport havi (ismét nominális) keresetét mutatja be, ahol egy már folyó programhoz kreáltak utólag véletlenszerűen leválasztott kontroll csoportot. A 0. hónap jelzi a program kezdetét, és a kritériumnak a kezdetkor és az azt megelőző 6 hónapban kellett megfelelni.

2. ábra: **Ashenfelter-horpadás egy átképzéssel kombinált munkahely-támogatási program kísérleti kontroll csoportjában** (Heckman és Smith, 1998)



Az Ashenfelter-horpadás akkor probléma, ha általában a program hatásának felülbecsléséhez vezet, ha csak a kezelt csoportnál létezik, a kontrollnál nem. Nem kísérleti kontrolloknál akik nem feleltek meg a program kezdetekor a szükséges kritériumoknak, általában nem jelenik meg a horpadás.

A példa kedvéért tegyük fel, hogy meg tudjuk figyelni az egyének mindazon jellemzőit, amelyek a munkaerőpiaci sikert befolyásolják. Van azonban egy olyan komponens, ami ugyanolyan munkaerőpiaci értékű egyének esetében is különböző eredményeket produkálhat: nevezzük ezt szerencsének. Tegyük fel még ezen kívül, hogy a szerencse forgandó, és hosszútávon szisztematikusan nem részesít senkit előnyben (ez hihető, ha minden olyan dolgot kontrollálni tudunk, ami a hosszú távú sikert befolyásolja). Ekkor az azonos munkaerőpiaci értékű egyének közül a programban részt vevők éppen a kritérium által adott időszakban szerencsétlenekből állnak. Minthogy azonban a szerencse forgandó, a program kritériumainak megfelelő egyének eredménye (foglalkoztatási valószínűség, kereset) egy későbbi időpontban várhatóan jobban fog növekedni, mint az azonos munkaerőpiaci értékű, de a kritériumnak nem megfelelő egyének várható eredménye – a program nélkül is. A kritériumnak megfelelők ugyanis nem foglalkoztatottak az indulás pillanatában, de pozitív valószínűséggel azok lesznek később. A kritériumnak nem megfelelők viszont átlagos, vagy az átlagosnál éppen szerencsésebb helyzetben voltak induláskor, ezért számukra nem jobb, sőt, lehet hogy rosszabb eredmények várhatóak később. Egy egyszerű összehasonlítás, de még egy DID (difference-in-differences) összehasonlítás is felfelé torzítja a program várható hatását, hiszen például pozitív hatást mutatna ki akkor is, ha nem is lenne semmiféle program. Lényeges felismerni, hogy ez a probléma akkor is releváns, ha minden olyan dolgot meg tudunk figyelni az emberekről, ami az ő munkaerőpiaci értéküket meghatározza.

Egy másik jelentős probléma a nem kísérleti kontroll csoport kiválasztásánál az idő dimenzió megfelelő kezelése. Az aktív foglalkoztatási eszközök többsége jelentős időt vesz igénybe (a munkakeresést segítő programok az egyetlen kivétel). Amíg a programban részt vevők támogatott munkában vagy átképzésben részesülnek, nem munkanélküliek: nem keresnek aktívan munkát. Felmerülhet az az igény, hogy a kontroll csoport olyan egyénekből álljon, akik a program lezárultakor felelnek meg a kritériumoknak. A program résztvevői ugyanis annak lezárulta után lépnek ki ismét a munkaerőpiacra, ezért jogosnak tűnik az az elvárás, hogy olyan egyének eredményeivel hasonlítsuk össze az övéiket, akik a program lezárultakor hasonló helyzetben vannak. Minthogy időközben megváltozhatott a gazdasági környezet, az elhelyezkedési, kereseti lehetőségek is változhattak időközben.

Fontos felismerni azonban, hogy pont a gazdasági környezet megváltozása miatt a kritériumoknak a program lezárultakor megfelelő egyének különbözhetnek a programban részt vevőktől: ők a megváltozott környezetben felelnek meg a kritériumoknak (vesztették el állásukat, kerültek ki a munkaerőpiacra az iskolából, stb.). Nyilvánvaló a probléma például a pályakezdő fiatalok elhelyezkedését segítő programok esetében: a későbbi időpontban megfelelő egyének más kohorszhoz tartoznak, ami miatt sok olyan dologban különbözhetnek a program résztvevőitől, amelyeket nem lehet megfigyelni (az iskolai képzés minősége, a kohorsz mérete és összetétele, stb.). Van még egy ok, ami miatt nem szerencsés a program lezárultának időpontjára koncentrálni: az esetek többségében nem is határozható meg ilyen időpont. A programok általában egy ideig folyamatosan futnak, és hosszabb időintervallum alatt vesznek fel és bocsátanak ki résztvevőket.

A helyes tényellentétes kérdés az, hogy mi lett volna a programban részt vevőkkel akkor, ha nem vettek volna részt. Ezért a megfelelő kontroll csoport mindig az, amelynek tagjai hasonló helyzetűek, így többek között eleget tettek a program kritériumainak *annak kezdetekor*. Elvi szinten a döntéshez egyszerű: az induláskori kritériumok azok, amik számítanak.

Ez igen pesszimista következtetés levonására ad alkalmat általánosan futó programok esetében, amelyekben mindenki részt vehet, aki megfelel a kritériumoknak. A legrosszabb esetben ilyenkor nem sok egyénből lehet az előbbiek értelmében megfelelő kontroll csoportot létrehozni. Ha van is elegendő számú ilyen egyén, az a saját döntése (vagy alulinformáltsága) eredményeképpen nem résztvevő. Ilyen esetben minden kontroll csoport nyilván szelektált, és ennek figyelmen kívül hagyása torzíthatja a hatások becslését. Megfelelő szelekciós vagy switching modellek használatával ez a probléma elvileg kezelhető. Minden ilyen indentifikáció feltétele azonban olyan instrumentum, amely befolyásolja a részvételt, de közvetlenül nincs hatása az eredményekre. Márpedig ilyen instrumentumokat általában nem könnyű találni.

4.2 Ökonometriai modellek nem kísérleti kontrollok esetén: LaLonde (1986)

Ebben az alfejezetben LaLonde (1986) nagy hatású tanulmányát elemezzük. A szakma igen ellentmondásos következtetéseket vont le ebből a cikkből: elsősorban az ökonometriai módszerek sikerességének (sikertelenségének) a kérdésére koncentráltak. Mint látni fogjuk azonban, sokkal inkább a megfelelő kontroll csoport kiválasztása a releváns kérdés, és LaLonde eredményei is csak ennek tükrében értelmezhetők.

A tanulmányban egy kontrollált kísérlet keretében megvalósított foglalkoztatáspolitikai programnak a keresetekre gyakorolt hatását vizsgálja. Abból indul ki, hogy a kontrollált kísérletből identifikálható a valós hatás. E hatás becslése után megvizsgálja, hogy milyen hatást lehetne becsülni akkor, ha a részt vevők eredményeit nem a kísérlet kontroll csoportjának eredményeivel hasonlítanánk össze, hanem más, a kutatók által gyakran használt kontrollcsoportokból becsült eredményekkel. Több nem kísérleti kontroll csoport eredményeit vizsgálja, és a hatás becsléséhez több különböző módszert is használ.

A LaLonde által vizsgált program az egyesült államokbeli National Supported Work (NSW) kísérlet volt. Az 1970-es évek közepén folytatott program kísérleti jellegű volt, és egy nonprofit szervezet végezte. A program célja az volt, hogy különösen hátrányos helyzetű embereket (akik a rendszeres munkához szükséges legalapvetőbb készségeknek sem voltak feltétlenül a birtokában) vezessen be (vagy vissza) a munka világába. A célcsoportba AFDC segélyben részesülő nők[13], volt kábítószerfüggők, volt elítéltek, és a gimnáziumból kibukottak tartoztak[14]. A programban való részvétel kritériuma minden esetben tartalmazta a valamilyen értelemben hosszabb távú munkanélküli státust.

A kísérletre jelentkezőket véletlenszerűen osztották kezelt és kontroll csoportba. A résztvevők (a kezelt csoport tagjai) számára a program 9-18 hónapi munkát garantált (célcsoporttól és attól függően, hogy melyik városban voltak). A résztvevőket néhány fős csoportokba osztották, és ezeket privát munkaadókkal alkalmaztatták őket támogatás fejében (minden munkaerőköltségüket a program fedezte). A program szervezői végigkísérték a résztvevő csoportokat, és rendszeres konzultációk során közösen megbeszélték problémáikat és teljesítményüket. A fizetett bér eleinte alacsonyabb volt az adott munkahelyen szokásosnál, de a munkavégzés rendszerességétől és a teljesítménytől függően később emelkedett. A résztvevők a garantált időszak végéig az adott munkahelyen maradhattak, de aztán rendes munkahelyet kellett találniuk (vagy rendesen foglalkoztatta őket a programban résztvevő munkaadó, vagy másutt kellet munkát keresniük). A

[13] Az AFDC program gyermekeiket egyedül nevelő nők számára nyújt élelmiszer támogatást Amerikában. Az NSW célcsoportba azok az AFDC támogatásban részesülők tartoztak, akik a program kezdeti időpontjában munkanélküliek voltak, legfeljebb 3 hónapot dolgoztak az azt megelőző fél évben, nem volt 6 évesnél fiatalabb gyermekük, és a megelőző 3 évben legalább 30 hónapban kaptak támogatást.

[14] Az Egyesült Államokban 18 évig tart az iskolakötelezettség; addig egységesen mindenkinek gimnáziumba (high school) kell járnia. A gimnáziumból kibukottak tehát vagy nem teljesítették az iskolakötelezettséget, vagy nem jutottak el annak utolsó évéig 18 éves korukra (utóbbi a ritkább esemény).

programban részt vevő nőket tipikusan szolgáltatási, a férfiak tipikusan építőipari munkahelyeken foglalkoztatták.

A kísérlet kezdetekor mindenkiről gyűjtöttek demográfiai és kereseti adatokat. A rendelkezésre álló adatok igazolják a hozzárendelés véletlenszerűségét: a kezelt és a kontroll csoport átlagos ismérvei kezdetben gyakorlatilag azonosak voltak. A szervezők a program alatt és egy ideig utána is gyűjtöttek adatokat a résztvevőkről és a kontroll csoport tagjairól is.[15] A program lezárulta után néhány évvel az AFDC támogatott nők között a részt vevők éves keresete átlagosan 850 dollárral haladta meg a kontroll csoport átlagos keresetét. A férfiak esetében nagyjából ugyanakkora, 880 dollár volt a különbség. A kis minták miatt azonban e különbségek meglehetősen imprecíz becslések: a nők esetében 300, a férfiakéban 500 dollár körüli a standard hiba.

A tanulmány ezeket a kísérletből mért hatásokat hasonlítja össze különféle nem kísérleti becslésekkel. Pontosabban fogalmazva az átlagos hatás becslése a cél, és a programban részt vevők tényellentétes eredményét (vagyis a programban való részvétel nélkül várható keresetét) becsüli meg több, nem kísérleti kontroll csoporton és többféle módszerrel. Ezeket hasonlítja aztán össze a kísérleti kontroll csoport eredményeiből becsült (fent részletezett) értékekkel. Fontos megjegyezni, hogy itt a nem kísérleti kiválasztású kontroll csoportok megfelelő összehasonlítható információ esetén elvileg használhatók a tényellentétes eredmények becslésére, mert a program kis méretű volt, és a kísérletben való részvétel kritériumainak megfelelő népességnek csak kis részére terjedt ki. Ezért elvileg találhatók olyan egyének, akik minden releváns szempontból hasonlítanak a programban részt vevőkre, de nem vettek részt a programban.

LaLonde két adatforrást használ (PSID és CPS-SSA összekapcsolt minta), amelyek országosan reprezentatívak, és amelyeket gyakran használnak (használtak) kontroll csoportként ilyen elemzésekhez. Ezekből, illetve ezeknek olyan almintáiból becsül különböző módszerekkel, amelyek tagjai eleget tesznek a kísérlet részvételi kritériumai egy részének (nem foglalkoztatottak a program kezdetekor, AFDC támogatásban részesülő nők).

Ezekből az adatforrásokból nem lehetett kialakítani az összes kritériumnak eleget tevők kontroll csoportokat, két okból sem. Egyrészt az adatok nem tartalmaztak minden relaváns információt a kritériumok meghatározásához. Másrészt, a beazonosítható kritériumoknak is mindössze néhány egyén fe-

[15] Későbbi interjúk nem mindenkivel készültek: az anyagi lehetőségek szűkülése miatt a szervezők egy mintát követtek csak végig, és ezek közül sem tudtak mindenkit elérni. A véletlen minta 65-75 %-os volt, ezen belül a lemorzsolódás 20-30 %-os volt.

lelt meg, így statisztikai elemzésre alkalmatlan minták jöttek volna ki belőle. Nem meglepő ezek után, hogy még a szűkített minták összetétele mérhető tulajdonságokban is nagyban eltér a kísérletben részt vevők (kezelt vagy kontroll csoportok) összetételétől. Ezen felül elvileg a nem kísérleti kontroll csoportban lehetnek olyanok is, akik valójában részt vettek a programban. Az országosan reprezentatív adatfelvételek ugyanis nem tartalmazták azt az információt, hogy a megfigyelt egyének részt vettek-e az NSW programban. Minthogy azonban a program rendkívül kis méretű volt, ennek a torzító hatása várható értékben elhanyagolható.

A vizsgált becslési módszerek a következők voltak: (1) a program lezárulta utáni keresetek egyszerű összehasonlítása; (2) a kor, iskola, és etnikum hatásának kiszűrése utáni összehasonlítás; (3) a program előtti és utáni keresetek különbségének (a keresetek növekedésének) összehasonlítása (DID, vagyis difference-in-differencess becslés); (4) a keresetek növekedésének különbségéből (DID) kiszűrve a kor hatása; (5) a program lezárulta utáni keresetek összehasonlítása a program előtti keresetek hatásának kiszűrése után; (6) a kezdeti keresetek mellett a kor, iskola, és etnikum hatásának kiszűrése utáni összehasonlítás; (7) a kezdeti keresetek mellett minden más mérhető változó hatásának kiszűrése utáni összehasonlítás.[16] Ezek lényegében mind keresztmetszeti regressziós modellek. LaLonde emellett négy különböző szelekciós modellt is vizsgál (ezek a használt instrumentumokban különböznek).[17] A szelekciós modelleket LaLonde csak a teljes mintákra becsüli meg, a részvételi kritériumokkal szűkített almintákra nem.

---

[16] Ezek a módszerek lényegében egyszerű átlag-összehasonlítások és keresztmetszeti regressziók. Y az eredményváltozó; X = kor, iskola, etnikum jobboldali változók ; Z az utolsó modellbe bevont összes többi változó (foglalkoztatás program kezdetekor, lakóhely, családi állapot, stb.); D pedig a programban való részvétel bináris változója (1 ha részt vevő, 0 ha kontroll). t index jelöli a program utáni mérést, t-1 a program előttit. m(Y) jelöli a mintaátlagot, az 1 index a kezelt csoportot, a 0 pedig a kontrollt. Jelölje δ a hatás becslőfüggvényét. A különböző modellekben ez a következőképpen definiálható (megj.: LaLonde nem írja le formálisan a modellek többségét; ezek a szövegből következtethetők ki):

(1) $Y_{it} = \delta D_i + U_i$ {*másképpen*: $\delta = m(Y_{1t}) - m(Y_{0t})$}

(2) $Y_{it} = \delta D_i + \beta X_i + U_i$

(3) $Y_{it} - Y_{it-1} = \delta D_i + U_i$ {*másképpen*: $\delta = m[(Y_{1t} - Y_{1t-1}) - (Y_{0t} - Y_{0t-1})]$}

(4) $Y_{it} - Y_{it-1} = \delta D_i + \beta age_i + U_i$

(5) $Y_{it} = \rho Y_{it-1} + \delta D_i + U_i$

(6) $Y_{it} = \rho Y_{it-1} + \delta D_i + \beta X_i + U_i$

(7) $Y_{it} = \rho Y_{it-1} + \delta D_i + \beta X_i + \gamma Z_i + U_i$

(5) – (7) tulajdonképpen kvázi DID becslőfüggvények, ahol a t-1 eredmény nem feltétlenül 1 együtthatóval vonódik ki a t-ik eredményváltozóból (kvázi-differencia differencia helyett).

[17] A szelekciós modell a következő volt (jelölések megegyeznek az előző lábjegyzetéivel):

Ha a tervezett kísérlet kontroll csoportjának tagjaival hasonlítjuk össze a programban résztvevőket, a fenti ökonometriai modellek szinte pontosan ugyanazt a hatást becsülik. Ez nem meglepő: ha a kísérletben a részvétel valóban véletlen kiválasztás eredménye volt, akkor ezt az eredményt kell kapnunk. A nem kísérleti kontroll csoportok esetében mintánként és módszerenként viszont egészen eltérő a hatások becslése. Önmagában ez sem meglepő: a különböző modellek különböző feltevéseken alapulnak, és a különböző alminták között is fundamentális eltérések vannak a kezelt csoporttal való összehasonlíthatóság tekintetében. A szelekciós modellek viszonylag jól szerepelnek, különösen az, amely instrumentumként a kezdeti foglalkoztatást és a lakóhelyet is tartalmazza. Ez szignifikánsan nem becsül eltérő hatást a két mintában és egyik sem tér el szignifikánsan a kísérleti becslésektől, ami azonban a magas standard hibáknak is köszönhető. A keresztmetszeti regressziós eredmények azonban igencsak problematikusak, mert esetükben még a legáltalánosabb, vagyis legkevesebb feltevésre építő modellek is csúnyán mellétrafálnak. A férfiak esetén – mintától függően – ráadásul nulla vagy negatív hatást becsülnek. Még egy problémát jelent az, hogy az őkonometriai modell-szelekciós módszerek egyáltalán nem a kísérleti eredményekkel összhangan levő modelleket választják ki. Kísérleti eredmények hiányában ezért nem lenne megmondható, hogy melyek a jobb modellek – illetve egyáltalán nem a legjobb modelleket választanánk ki.

LaLonde (1986) tanulmányának rendkívül nagy hatása volt a társadalompolitikai programok hatásvizsgálatának gyakorlatára. Minthogy a cikk nagy része a különböző becslési eljárások összehasonlításával foglalkozik, a szakma egy része azt a következtetést szűrte le az eredményekből, hogy az ökonometriai módszerek – talán a szelekciós modellek kivételével – nem sok mindenre használhatók. Ez az ítélet azonban elhamarkodott. LaLonde vizsgálatában a valós probléma nem az, hogy az ökonometriai modellek *általában* nem identifikálják a mérni kívánt hatást, hanem hogy *az általa vizsgált mintákban és ott mérhető változók használatával* nem identifikálják azt.

Megalapozottabb interpretációk (például Heckman, Ichimura és Todd, 1997; Heckman, Ichimura és Todd, 1998) szerint a nem kísérleti modellek

---

$Y_{it} = \rho_1 Y_{it-1} + \beta_1 X_i + \gamma_1 Z^*_i + U_i$

$d_i = \rho_2 Y_{it-1} + \beta_2 X_i + \gamma_2 Z_i + V_i$

d_i a programban való részvételt indukáló látens változó (D_i=1 ha d_i>0, D_i=0 egyébként). Z mindazon változók amelyek az előző lábjegyzetben szerepeltek; ezek közül legfontosabb a foglalkoztatás program kezdetekor. Z^* Z változóinak egy részét tartalmazza. A LaLonde által becsült négy specifikáció abban különbözik, hogy Z^* a Z vektor mennyi változóját (és melyeket) tartalmazza (az egyik extrém specifikációban Z^*=0, a másikban Z^*=Z).

kudarca valójában a nem kísérleti kontroll csoportok kiválasztásában és fontos adatok hiányában keresendő. A kontrollként felhasznált adatokat nem az adott program hatásának mérésére gyűjtötték, hanem országosan reprezentatív minták voltak. Legfontosabb problémáik a következők voltak: nem a megfelelő helyi munkaerőpiacról választották ki a válaszadókat, nem mindig volta összehasonlíthatók a változók, és a kontroll csoportokban nem érvényesítették a programban való részvétel kritériumait. A kritériumok egy része nem is volt beazonosítható ezekben a mintákban, így nem is lehetett kompatibilis almintát létrehozni. Még ha lehetett volna is, az olyan kicsi lett volna, amely a meghiúsította volna a tárgyalt ökonometriai módszerek használatát.

LaLonde (1986) tanulmányának fő tanulsága tehát az, hogy *a nem kísérleti kontroll minták kiválasztása és az összehasonlításhoz fontos információk megléte valójában fontosabb, mint a program hatásának becslésére használt mérési módszer*. A módszerekkel kapcsolatban is van azonban mondanivalója: a szelekciós modellek – ha megfelelő instrumentumokat tudunk találni – jobbak lehetnek az egyszerű keresztmetszeti modelleknél. Ez a következtetés nagymértékben függ a felhasználható változóktól: a vizsgált adatokban például hiányzik a programot megelőző, hosszú távra visszamenő munkaerőpiaci történet. Az azt is felhasználó modellek (amelyek persze már csak formálisan maradnak keresztmetszeti regressziók, valójában az idő dimenziót is kiaknázzák) sokkal jobban szoktak teljesíteni – ezt azonban itt nem lehet lemérni. LaLonde nem vizsgálta a párosítási modelleket sem, mivel azokat a nyolcvanas évek első felében nem használták széles körben hatásvizsgálatokra.

4.3 PÁROSÍTÁSI MODELLEK NEM KÍSÉRLETI KONTROLLOK ESETÉN

Dehejia és Wahba (1998, 2003), valamint Smith és Todd (2002) újraelemezték a LaLonde által vizsgált NSW az adatokat, és párosítási (matching) módszerekkel újrabecsülték a nem kísérleti kontroll csoportokkal mért hatásokat. Dehejia és Wahba (1998) LaLonde teljes (adathiánytól eltérő okok miatt nem szelektált) nem kísérleti mintáin párosítási modellekkel becsülte a program hatását, és LaLonde-nál sokkal kedvezőbb eredményeket kapott.

A szerzők rétegzett (stratified) propensity score becslést alkalmaztak. Ennek lényege, hogy a résztvevő és a nem kísérleti kontroll csoportokat a becsült propensity score diszjunkt intervallumai alapján almintákba osztották, és ezeken belül végezték aztán el az összehasonlítást.[18] Ha a kontrollált

[18] A propensity score becslés parametrikusan történt: a magyarázó változók felsőbb hatványait és interakcióit is tartalmazó, tehát meglehetősen flexibilis logit modellel. Az összehasonlítást a propensity score becsléshez is használt regresszorok hatását ki-

magyarázó változókon kívül más nem befolyásolja a részvételt (legalábbis nem olyan, ami korrelál az eredménnyel), akkor az intervallumokon belüli eredmények súlyozott átlaga torzítatlan becslést ad a program hatására. [19] Dehejia és Wahba (1998) eredményei alapján a nem kísérleti kontroll csoportokon párosítási modellekkel becsült hatások nagyon közel vannak a kísérleti kontroll csoport összehasonlításával becsült hatáshoz. Emlékeztetőül: LaLonde ugyanezeken a nem kísérleti kontroll csoportokon ökonometriai modellekkel becsült hatásokat, amelyek – talán a szelekciós modellek kivételével – a kísérleti eredményekkel köszönő viszonyban sem voltak. Első ránézésre tehát az a következtetés látszik helyesnek, hogy a párosítási modellek a keresztmetszeti regresszióknál és DID becsléseknél jobb eredményeket adnak.

Valójában azonban megint nem a becslési módszerek különbsége a magyarázat az eltérő eredményekre. A párosítási modellek csupán explicit módon kezelnek olyan problémákat, amelyek hagyományos ökonometriai modellek esetében a becslés előtt, a kontroll csoport kiválasztásánál merülnek fel. Dehejia és Wahba a két országosan reprezentatív minta túlnyomó részét (közel 80 százalékát) nem tudta a kezelt csoporthoz párosítani, mert esetükben a kezelteknél becsült minimum alatti propensity score-t becsült. A maradék csaknem 90 százaléka pedig a legalacsonyabb (0.05 alatti) propensity score kategóriába került – a kezelt csoport kevesebb, mint 10 százaléka esett ide. Magyarul a nem kísérleti kontroll csoportokból (emlékeztetőül, ezek az országosan reprezentatív PSID és CPS minták) nagyon

---

szűrve is elvégezték, ám ez gyakorlatilag ugyanazt adta, mint az egyszerű különbség. Rosenbaum és Rubin (1983) fenti 2. tétele alapján ezt várnánk, ha a propensity score-t nemparaméteresen becsülték volna. Az eredmény itt azt jelenti, hogy a flexibilis logit specifikáció elég jól közelítette a nemparaméteres eredményeket. Az eredmények egyébként nem változtak akkor sem, amikor a logit modellben nem szerepeltették a változók felsőbb hatványait és interakcióit.

[19] A becslési algoritmust Dehejia és Wahba (2002) részletezi. Ennek lépései a következők:

- Viszonylag egyszerű logit specifikáció alapján a propensity score becslése
- Az adatok rendezése a becsült propensity score alapján
- A megfigyelések besorolása a propensity score vizsonylag kicsi diszjunkt intervallumai (pl. 0–0.02, 0.02–0.04, stb) alapján.
- Megvizsgálni, hogy a megfigyelhető változók (X) azonos eloszlásúak-e az egyes intervallumokon belül a kezelt és a kontroll csoportban.
- Ha azonos eloszlásúak, az eredmények összehasonlíthatóak, és a hatás becslése elvégezhető (kondicionálva a propensity score-ra, és átlagosan is).
- Ha nem azonosak, két dolgot lehet tenni: finomítani az intervallum-beosztást, és flexibilisebbé tenni az első lépés logit specifikációját. Ezeket addig kell variálni, amíg azonos eloszláshoz nem jutunk. Akkor elvégezhető a hatás becslése (lásd előző lépés).

kevesen (az eredeti minták 2-3 százaléka) rendelkeztek olyan megfigyelhető ismérvekkel, amelyek az elhanyagolhatónál nagyobb valószínűséggel tette volna őket a program résztvevőivé. Ez természetesen nem meglepő, hiszen a program különösen hátrányos helyzetű csoportokat célzott meg, a kontroll csoportok viszont országosan reprezentatív minták voltak. Ha a propensity score becsléshez felhasznált változók elég sok információt tartalmaztak, akkor a becslés eredményeként a nem kísérleti mintákból is csak a leginkább hátrányos helyzetűek voltak párosíthatók. Bár – mint azt fent már említettük – a program összes kritériumáról nem tartalmaztak információt az országosan reprezentatív minták, a jelek szerint elég megfigyelhető ismérv volt bennük ahhoz, hogy a párosítási arányok, és azután a becsült hatások is jól közelítsék a valóságot.

Dehejia és Wahba tehát a becsléshez a nem kísérleti kontroll csoportnak csak egészen kis részét használták fel. A párosítási modellek nagy előnye, hogy a kontroll csoportnak ezt a szelekcióját explicit módon megteszi. Elvileg azonban mindez megtehető lenne a hagyományos ökonometriai módszerek esetében is, csak akkor mindennek a becslést megelőzően, a kontroll csoport kiválasztásánál kellene megtörténnie. A LaLonde (1986) tanulmány következtetéseit ért legfontosabb kritikák pont erre mutattak rá: ott nem történt meg ilyen előzetes szelekció, illetve nem elég körültekintően történt meg. Ez utóbbin természetesen azt kell érteni, hogy LaLonde a részvételi kritériumoknak elvileg jobban megfelelő alminták kiválasztásánál nem vett figyelembe minden információt – ezen információk egy része nem volt hozzáférhető, de a kritikák szerint fontos hozzáférhető információk is kimaradtak.

Dehejia és Wahba (1998) nem túl explicit a tekintetben, hogy pontosan milyen mérhető változókat (X) használtak a propensity score becsléshez. Érzékenységi vizsgálatukban a legfontosabbnak ítélt változó, a program kezdete előtti kereset szerepeltetését vizsgálják. Az eredmények azt mutatják, hogy annak kihagyása nagymértékben torzítja a becsült hatást, ám azok még mindig közelebb vannak a kísérleti összehasonlításból identifikált hatáshoz, mint a LaLonde által ökonometriai modellekkel becsült hatások (ismét a szelekciós modellek egy részének eredményeit leszámítva).

Smith és Todd (2002) elemzései azt mutatják, hogy Dehejia és Wahba (1998) eredményeit nagyban meghatározza az, hogy valójában nem a teljes részvételi mintára becsülik a párosítási modelljeiket. A kezelt mintát ugyanis leszűkítették azokra az egyénekre, akikre rendelkezésre állt kereseti adat a program kezdete előtt 2 évvel. Erre a kritériumok előtti keresetek kontrollálása miatt volt szükség, ám így a kezelt csoport mintegy 40 százalékát kizárták az elemzésből. A mintaszelekció nem volt véletlenszerű,

amelynek egyik következménye az lett, hogy a kísérleti összehasonlításból becsült hatás csaknem kétszer akkora, mintha a teljes résztvevő csoportra becsülnénk. A mintaszelekció másik következménye az volt, hogy a maradék kezelt csoportban az Ashenfelter-horpadás jóval kisebb lett, mint a teljes mintában. Az eredeti Dehejia és Wahba-féle kezelt csoport mintát némiképp módosítva Smith és Todd bemutatták, hogy erre a kezelt csoportra a LaLonde (1986) által vizsgált ökonometriai modellek is sokkal jobb eredményeket hoznak a nem kísérleti kontroll csoportokon. Az egyszerű DID becslések például még közelebb vannak a kísérleti eredményekhez, mint a párosítási modellek. A Dehejia és Wahba (1998) által becsült propensity score alapú párosítási modelleket a kezelt csoport más mintáira alkalmazva pedig ugyanúgy torz eredményeket kaptak, mint ökonometriai modellekkel. Azt a következtetést vonják le, hogy a párosítási modellek önmagukban egyáltalán nem jelentenek gyógyírt a nem kísérleti kontroll csoportokkal való összehasonlítás esetén.

Viszontválaszában Dehejia (2003) rámutat arra, hogy a Smith és Todd (2002) által elvégzett vizsgálatok eredményei félrevezetőek. A Dehejia és Wahba által becsült propensity score modellek speciálisan az általuk kiválasztott kontroll csoportokra értelmesek. Más és egymástól is eltérő nem kísérleti kontroll csoportokra ezeknek a specifikációknak nem feltétlenül van értelmük, így nem meglepő, hogy azokra torz eredményeket adnak. Dehejia érvei szerint kezelt csoport általuk vizsgált almintája felel meg leginkább a hatásvizsgálatok egyik alapvető követelményének: annak, hogy a program kezdete előtt megfelelő távolságban (itt két év) legyenek adatok az egyének eredményváltozójáról (itt keresetéről). Ez a követelményt Ashenfelter (1978) és Ashenfelter és Card (1985) mellett Heckman, Ichimura, Smith és Todd (1997) is hangsúlyozta.

A vita legfontosabb tanulsága itt is az, hogy a nem kísérleti kontroll csoportok kiválasztása és a szükséges információk megléte, ha lehet, fontosabb az alkalmazott becslési modellnél. Pontosabban fogalmazva, a kettő nem választható el egymástól. A párosítási modellek elvitathatatlan előnye az, hogy ezeket a kérdéseket explicit módon vetik fel.

4.4 Általános következtetések

Heckman, Ichimura, Smith és Todd (1997) különböző nem-kísérleti kontroll csoportokon becsült párosítási modellek torzításait vizsgálja egy másik aktív foglalkoztatási program esetén, és hasonlítja össze azok eredményeit tervezett kísérleti becslésekkel. A legfontosabb következtetésük az, hogy a legfontosabb az adatok minősége és összehasonlíthatósága. Esettanulmányukban azt a következtetést vonták le, hogy a nem kísérleti becslések ak-

kor eredményeztek a kísérleti eredményekhez közeli eredményeket, ha a következő három feltétel teljesült:

- A résztvevők és a kontrollok adatai minden részletükben összehasonlíthatók voltak. Ez a követelmény automatikusan teljesül azonos forrású adatok esetén, más esetekben azonban igen nehéz érvényt szerezni neki;
- Az adatok sok olyan változót tartalmaztak, amelyek relevánsak voltak a részvételi döntés modellezésében (a program előtti, több évre visszamenő munkatörténet és keresetek a legfontosabbak);
- A résztvevők és a kontrollok azonos munkaerőpiacon éltek.

Bár a tanulságok egyetlen esettanulmányon alapulnak, általános érvényű követelményként kezdenek elterjedni a nem kísérleti hatásvizsgálatok esetére. A sok megfigyelhető változó követelménye meglehetősen evidens, és a hosszú időre visszanyúló munkatörténet szükségessége szinte minden más elemzés esetén is beigazolódott. A legelső követelményt azonban értelmezik úgy is, hogy az azonos adatforrás szükséges feltétele az elemzésnek. Ez nyilvánvalóan téves következtetés, hiszen az összehasonlíthatóság az, ami szükséges, és az azonos forrás ennek csupán elégséges feltétele. Amennyiben vakon elfogadjuk a fenti hármas követelményrendszert, akkor például a hozzárendelés területi szóródásából identifikáló természetes kísérleti megközelítés – a kezelt csoport az a munkaerőpiac, ahol bevezették a programot, a kontroll csoport az, ahol nem – eleve ki van zárva. Nyilvánvaló, hogy különböző munkaerőpiacokon élők eredményeinek összehasonlítása nehéz, hiszen a különböző helyeken eltérő dolgok történhettek a vizsgálat ideje alatt. Ennek kontrollálása azonban megfelelő információk birtokában elvileg elvégezhető. Az azonos munkaerőpiacon élés elégséges, ám megint nem szükséges feltétele az összehasonlíthatóságnak. Az összehasonlíthatóság elvét kell szem előtt tartani, nem pedig azok elégséges feltételeit. A probléma kezelése korántsem egyszerű, de amint Dehejia és Wahba (1998) és Blundell, Costa Dias, Meghir és Van Reenen (1997) később elemzett tanulmányai is mutatják, nem reménytelen.

5. A HATÁSOK MÖGÖTTI MECHANIZMUSOK MÉRÉSÉNEK PROBLÉMÁJA

A program átlagos hatásának identifikálása mellett nyilvánvalóan fontos kérdés az, hogy miért és hogyan érvényesül a hatás – már ha létezik egyáltalán. Ebben a fejezetben a hatások mögötti mechanizmusok mérésének a problémáját vizsgáljuk meg. Először röviden bemutatjuk a mérési probléma gyökereit, általános szinten. Ezután egy konkrét problémán keresztül is

megvizsgáljuk a kérdést és az arra adható válaszokat. A példa a már bemutatott NSW tervezett kísérletre épül, ahol az aktív munkaerőpiaci programnak van kimutatható pozitív hatása a foglalkoztatási esélyekre, ám e hatás mögött alternatív mechanizmusok állhatnak. A példa Ham és LaLonde (1996) tanulmánya.

5.1 A MECHANIZMUSOK MÉRÉSÉNEK ÁLTALÁNOS PROBLÉMÁJA

A probléma általánosan (bár kissé bonyolultan) megfogalmazva a következő. Ha a mechanizmusok méréséhez olyan kimenetelt kell vizsgálni, amelyek olyan állapothoz köthetőek, amelyek előfordulását a program maga befolyásolja, akkor akár kontrollált kísérletről van szó, akár nem, az ebben az állapotban levőket szelektálja maga a program hatása. E szelekció azt eredményezi, hogy a kezelt és a kontroll csoport összetétele – mérhető és nem mérhető ismérveik alapján – különbözik a két állapotban, még akkor is, ha az eredeti hozzárendelés véletlenszerű volt. A csoportok átlagos kimeneteleinek egyszerű összehasonlítása így nem identifikálja a mechanizmusokat feltáró kimenetelekre gyakorolt átlagos hatást, még akkor sem, ha a kezelt és a kontroll csoportba tartozás véletlen kiválasztás eredménye volt.

Tegyük fel, hogy a program hatása pozitív Y-ra, és ez egy W változón keresztül érvényesül. A hatás mögötti mechanizmus feltárása tehát nem jelent mást, mint megmérni, a programnak milyen hatása volt W–re. E hatás felmérésének a problémája azonos az eredeti identifikációs problémával:

$W_{1i} \mid D_i = 1$ mérhető, de $W_{0i} \mid D_i = 1$ nem mérhető; és

$W_{1i} \mid D_i = 0$ nem mérhető, de $W_{0i} \mid D_i = 0$ mérhető.

A hatások mögötti mechanizmusok mérésének a problémája valójában nem különbözik a hatások mérésének a problémájától: mindössze egy másik eredményváltozót vizsgálunk. Akkor állunk szemben speciális problémával, ha Y korlátozza W megfigyelhetőségét. Legyen a program hatását mérő kimenetel változó bináris (foglalkoztatás):

$Y_i = 0$ vagy 1.

Amennyiben W csak akkor figyelhető meg, ha Y=1, akkor nemvéletlen szelekciós probléma léphet fel D_i véletlen hozzárendelése esetén is.

Tekintsük a következő példát (ez a Ham és LaLonde, 1996, későbbiekben ismertetett tanulmányában is előforduló egyik probléma stilizált változata). Legyen Y a foglalkoztatást jelző bináris változó, W pedig az állásban eltöltött idő. Nyilvánvaló, hogy azok, akik tovább képesek megtartani az állásu-

kat, bármely adott időpillanatban nagyobb valószínűséggel lesznek foglalkoztatva:

$E(Y_i \mid W_i) = P(Y_i=1 \mid W_i)$ pozitív függvénye W_i-nek.

A kérdésünk az, hogy a program átlagos pozitív hatása Y-ra (a foglalkoztatási valószínűségre) W-n keresztül érvényesül-e. Vagyis, hogy a program hatására megváltozott-e valami a résztvevők legalább egy részével kapcsolatban, amitől azok tovább maradnak állásban (pl. elsajátítottak egy foglalkozást, vagy "belerázódtak" a munka világába).

Példánkhoz tegyük fel azt is, hogy az állásban eltöltött idő függ egy, a programtól független változótól, V-től (pl. alapvető motiváltság vagy olyan képességek, amit a programban való részvétel nem változtat meg).

$E(W_i \mid V_i)$ pozitív függvénye V_i-nek, így

$E(Y_i \mid V_i) = P(Y_i=1 \mid V_i)$ pozitív függvénye V_i-nek is.

Kérdésünk az, hogy a program – ceteris paribus, tehát V adott értékei mellett – növeli-e W-t várható értékben:

$E(W_{1i} \mid V_i, D_i=1) - E(W_{1i} \mid V_i, D_i=0) = ?$ és

$E(W_{0i} \mid V_i, D_i=1) - E(W_{0i} \mid V_i, D_i=0) = ?$

Amennyiben a kezelt és a kontroll csoport V szerinti összetétele nem ugyanaz az Y=1 állapotban, akkor két csoport állásban eltöltött idejének egyszerű különbsége nem identifikálja a mérni kívánt átlagos hatást. Eddig ez az eredeti identifikációs probléma újrafogalmazása egy másik eredményváltozóra (W-re). Ha a csoportokba rendelés véletlenszerű, akkor V alapján sem különbözik a két csoport (várható értékben):

$E(V_i \mid D_i=1) = E(V_i \mid D_i=0)$.

A plusz identifikációs probléma abból fakad, hogy W csak Y=1 esetén megfigyelhető. Tehát ami valójában mérhető, az az Y=1 állapotúak közötti különbség:

$E(W_i \mid D_i=1, Y_i=1) - E(W_i \mid D_i=0, Y_i=1)$.

Ekkor ugyanis, ha a programnak van pozitív hatása Y-ra, a két csoport azon része, ahol Y=1 (akik foglalkoztatottak), már nem lesz azonos összetételű V alapján. Ennek az az oka, hogy ha a program hatása pozitív akármilyen okból kifolyólag, akkor a kezelt csoportban többen foglalkoztatottak. Így viszont a kezelt csoportban lesznek olyan, alacsonyabb V értékű egyének is, akik a kontroll csoportban nem lennének foglalkoztatva. Így a kezelt csoport foglalkoztattjai között átlagosan alacsonyabb lesz V:

$E(V_i \mid D_i=1, Y_i=1) < E(V_i \mid D_i=0, Y_i=1)$.

Mivel pedig $E(W_i | V_i)$ pozitív függvénye V-nek, ceteris paribus alacsonyabb átlagos W-t (állásban töltött időt) fogunk mérni a kezelt csoport tagjainál. Egy ilyen ceteris paribus helyzet lehet az, amikor a foglalkoztatási arányt a program nem W-n, hanem valami más úton növeli meg. Ekkor effektive alacsonyabb állásban töltött időt fogunk mérni a kezelt csoportban a program után, ez azonban nem a program negatív következménye, mindössze egy szelekció eredménye.

Amennyiben V mérhető, egyszerűen kontrollálni kell rá. Azonos V-vel rendelkező kezelt, illetve kontroll megfigyeltek esetén a program hatására változó (vagy nem változó) W-t fogunk mérni. Valós mérési probléma akkor keletkezik, ha V nem mérhető (nem mérhető tökéletesen).

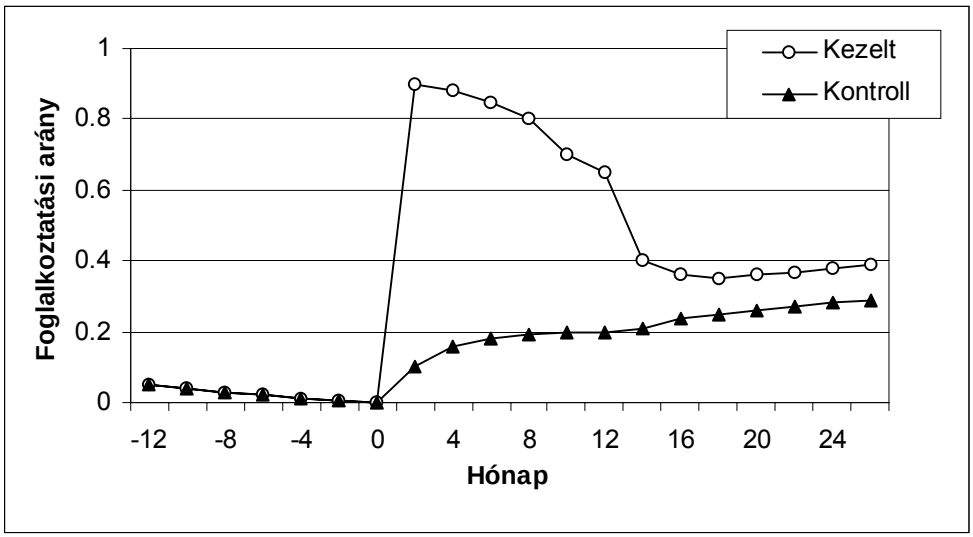
A következő alfejezetben egy példán keresztül bemutatjuk a probléma relevanciáját. Az esettanulmányban a kezelt, illetve kontroll csoporthoz rendelés véletlenszerű volt, így az Y-ra gyakorolt átlagos hatást egyszerű összehasonlítás identifikálja. A mért pozitív hatás mögötti okok feltárása azonban az itt vázolt identifikációs problémákba ütközik.

5.2 HAM ÉS LALONDE (1996)

A vizsgált program itt is az egyesült államokbeli National Supported Work (NSW) kísérlet volt, mint LaLonde (1986) tanulmányában. Ham és LaLonde ennek az AFDC támogatásban részesülő női almintáját vizsgálják. Emlékeztetőül: a programnak ez a része különösen hátrányos helyzetű, a kísérlet kezdetének idején nem (és tipikusan azt megelőzően sem) dolgozó nők számára biztosított 12 hónapnyi foglalkoztatást privát munkaadóknál, de minden munkaerő költségüket a program fedezte. Folyamatos konzultációkkal segítették a résztvevőket, akiknek a munkahelyen fokozatosan emelkedő elvárásoknak kellett megfelelniük. Ez tehát egy olyan támogatott foglalkoztatási program volt, amely munkahelyi ösztönzési és informális képzési elemekkel volt kombinálva. A kísérletre jelentkezőket véletlenszerűen osztották kezelt és kontroll csoportba. A vizsgált kimenetel ebben a tanulmányban a foglalkoztatás volt a program lezárulta (itt 12 hónap) után, amikor a munkaadóknak már nem volt kötelessége foglalkoztatni a programban résztvevőket. Emlékeztetőül: LaLonde a keresetet vizsgálta, ami a foglalkoztatás mellett azt is figyelembe vette, hogy aki dolgozott, az milyen sikerrel tette azt (a nem foglalkoztatottak keresetét ott 0-val vette figyelembe). Itt maga a foglalkoztatás az eredményváltozó. A véletlen kiválasztás eredményeként a két csoport összetétele minden megfigyelhető ismérv alapján azonos volt, így a programba kerülés előtti történet szerinti megoszlásban sem volt különbség.

A program hatása foglalkoztatásban mérve is pozitív volt: a kontroll csoportnál magasabb arányban voltak foglalkoztatottak a kezelt csoport tagjai a 12 hónap letelte után is. (Emlékeztetőül: a pozitív hatás a keresetekben is kimutatható volt.) A *3. ábra* mutatja be a két csoport foglalkoztatási arányait a program előtt, közben és utána.

3. ábra: **Az NSW kezelt és kontroll csoportjának foglalkoztatási aránya a kísérlet előtt** (-12-0. hónap)**, közben** (0-12. hónap) **és utána** (12-26. hónap)



Ham és LaLonde (1996) nyomán[20]

Az ábra önmagában is rengeteg információt tartalmaz. A program kezdete előtt a két csoport foglalkoztatása teljesen azonos (és minimális) volt. A program alatt a kezelt csoport tagjait támogatott munkahelyeken foglalkoztatták, ami induláskor 90%-os eredménnyel járt. Fokozatos lemorzsolódás után a támogatott szakasz végére, vagyis a 12. hónapra ez 65%-ra csökkent, majd a támogatás megszűntével 40%-ra esett. A kezelt csoport tagjainak foglalkoztatása azonban több mint egy évvel a program vége után is csaknem 10 százalékponttal magasabb volt, mint a kontroll csoport tagjaié. A kísérlet hosszú távú átlagos hatása tehát durván 10 százalékpontos (relatív terminusokban 30 %-os) foglalkoztatási esélynövekedésre becsülhető.

A fő eredmények mellett három furcsaság is megfigyelhető. Az egyik az, hogy a foglalkoztatási arány csak a program kezdetekor és az azt megelőző néhány hónapban volt 0, előtte és utána is annál magasabb volt (a kontroll csoportban is). Ez a jelenség a fentebb már tárgyalt "Ashenfelter-horpadás" (Ashenfelter's dip). Itt arról van szó, hogy a kísérletben való részvétel

[20] Az ábra a publikáltnak közelítőleges reprodukciója.

egyik feltétele az, hogy az egyén ne dolgozzon az indulás időpontjában és azt néhány hónappal megelőzően. Több hónappal előtte ilyen megkötés nincs, így a program szelektál az általában alacsony foglalkoztatottságúak közül: csak azok vehetnek benne részt, akik a kitüntetett időszakban egyáltalán nem dolgoztak. Az is természetes, hogy valamekkora foglalkoztatás-emelkedés a kontroll csoportban is megtalálható a 0. hónap után: a program hiányában is dolgozni fognak néhányan, mint ahogyan a nemfoglalkoztatottsági kritérium előtt is ez volt a helyzet. Tipikus esetben kísérleti kontroll csoportok esetében a "horpadás" szimmetrikus (a 2. ábrán is az volt), vagyis a növekedés a kontroll csoportban a korábbi szint elérése után megáll. Itt azonban nem.

Ez a második figyelemreméltó jelenség: az "Ashenfelter-horpadás" erősen aszimmetrikus. Esetünkben a kontroll csoportban is a korábbi szintnél jóval magasabbra ugrik a foglalkoztatás: egy év alatt 20 százalékosra. Ennek két magyarázata lehet. Működhetett egyféle "helyettesítési hatás", amikor a véletlenül a kontroll csoportba kerültek is valójában több segítséget kaptak a program beindulása előttinél. De lehetséges, hogy az általános gazdasági környezet javult meg annyira, hogy több mint kétszeresére nőtt volna a célcsoport foglalkoztatása a program nélkül is. Előbbi eset mindenképp probléma, és a program hatásának alulbecsléséhez vezet. Minthogy nincs semmiféle információnk a helyettesítési hatás meglétéről (és a különböző helyszínek közötti esetleges szóródásáról), ezzel a torzítással nem lehet mit kezdeni. A javuló gazdasági helyzet viszont nem feltétlenül jelent problémát, hiszen a kezelt és a kontroll csoport tagjainak ugyanaz az összetétele, és a program lezárulta után ugyanazon megváltozott munkaerőpiacon kell helytállniuk.

Mindezeken kívül az is megfigyelhető, hogy a program lezárulta (a 12. hónap) után a kezelt csoport foglalkoztatása leesett, majd tovább csökkent, és ezután kicsit megint emelkedni kezdett. A dolog hátterében az áll, hogy a támogatott foglalkoztatás megszűnte után a többséget nem vették át a munkáltatók saját költségükre, és ezek jelentős része nem talált azonnal más munkát. A résztvevő munkavállalók közül azonban néhánynak jól láthatóan sikerült később új munkát találnia.

A tanulmány fő kérdése az, hogy mi áll a csaknem 10 százalékpontos átlagos hatás mögött. A szerzők két lehetséges magyarázatot találtak: (1) A kezelés növelte a résztvevők termelékenységét, és így azok jobb munkaerővé váltak. Emiatt esetleg nagyobb eséllyel is veszik fel őket, az igazán fontos hatás azonban az, hogy ha már egyszer felvették őket valahova, nagyobb eséllyel tudják megtartani a munkahelyüket. (2) A kezelés a munkahelyre való felvételt segítette (azáltal pl. hogy a résztvevők jobban tudják, hogyan

gondolkodnak a felvételiztetők), ám nem volt hatással a munkahelyi termelékenységre, így a résztvevők munkahely-megtartó képességeire sem.

A két különböző mechanizmusnak eltérő jóléti és társadalompolitikai következményei lehetnek. Elképzelhető például, hogy az elhelyezkedési lehetőségeket hasonló mértékben lehet javítani jóval olcsóbban is egy ilyen támogatott foglalkoztatási programnál.

A kérdés eldöntése elvileg egyszerű: meg kell vizsgálnunk, hogy a kialakult foglalkoztatottsági különbség mögött mennyiben áll az, hogy kezelt csoport tagjai (1) hosszabb ideig maradnak egy munkahelyen ha felveszik őket, és mennyire az, hogy (2) hamarabb tudnak elhelyezkedni, vagyis rövidebb ideig munkanélküliek. A munkahelyen töltött időt ugyanis csak az első mechanizmus hosszabbítja meg. A rövidebb munkanélküliség pedig következménye kell hogy legyen a második mechanizmusnak. Két kimenetel átlagát kell tehát vizsgálni: az állásban maradás, illetve a munkanélküliség időtartamáét.

A kezelt és kontroll csoportba tartozás azonban hiába véletlenszerű, az időtartamok (a fenti jelöléssel W) egyszerű összehasonlítása nem identifikálja a mérni kívánt átlagos hatást, a szelekciós torzítás miatt. Feltételezhetjük ugyanis, hogy a programtól függetlenül jobb eredeti képességűek – a fenti jelölésekkel magasabb V-vel rendelkezők – tovább maradnak állásban. (Eredeti képességekről beszélünk, ami alatt a kísérlet előtti termelékenységet értjük – a kezelés hatására ugyanis a termelékenység változhatott, ám ebben nem lehetünk biztosak, hiszen éppen ennek vizsgálata a célunk.) Egészében a két csoport közel azonos összetételű (a véletlen kiválasztás miatt), így a teljes csoportokban az átlagos eredeti képességek is közel azonosak. Mivel azonban a kezelt csoportból nagyobb arányban vannak állásban, ott az eredetileg kicsivel rosszabb képességűek is állásba kerültek, szemben a kontroll csoporttal. A 12. hónap után állásban levők között így a kezelt csoportban az eredeti képességek átlaga alacsonyabb.

Ham és LaLonde megmutatja, hogy a program után foglalkoztatottak között a kezelt és kontroll csoport összetétele valóban eltér néhány fontos mérhető ismérvben. A legfontosabb ezek között az, hogy kezelt csoportból foglalkoztatottak kicsit kevésbé iskolázottak, és átlagosan jóval kevesebb program előtti munkaerőpiaci tapasztalattal rendelkeznek. Alacsonyabb iskolázottság és tapasztalat a munkaerőpiacon általában alacsonyabb munkahelyi termelékenységgel, és így kisebb munkahely-megtartási valószínűséggel jár.

Első megközelítésben szimmetrikus probléma a munkanélküliség időtartamának összehasonlítása. Ha a program hatása legalább részben a jobb

munkahely-megtartási képességben van, akkor a kezelt csoportból munkanélkülivé válók eredeti képességeik alapján átlagosan rosszabbak, mint a kontroll csoport munkanélküli tagjai. Ha alacsonyabb eredeti képességek hosszabb munkanélküli időtartammal járnak ceteris paribus, akkor a kezelt csoport tagjai hosszabb ideig maradnak munkanélküliek amennyiben a programnak nincs hatása az elhelyezkedési valószínűségre. Ha pedig van hatása arra is, akkor azt egyszerű összehasonlítás alulbecsüli a szelekciós torzítás miatt.

A munkanélküliség időtartam vizsgálatának azonban még egy külön nehézsége is van. A kezelt csoport tagjai ugyanis megszakították munkanélküli állapotukat a program idejére. A kontroll csoport tagjai viszont továbbra is munkanélküliek maradtak. A kontrollok munkanélküli időtartama már pusztán emiatt is hosszabb a kezeltekénél. Ráadásul azok a kontrollok, akik a program lezárulta után is munka nélkül voltak, a kontroll csoporton belül átlagosan rosszabb képességűek. Ennek egyik oka az, hogy a jobbak nagyobb valószínűséggel találtak munkát időközben – ezt az időtartamok vizsgálatánál a heterogenitás miatti szelekciónak nevezik, amely akkor probléma, ha nem megfigyelt heterogenitásról van szó. Másik oka az lehet, hogy a hosszabb munkanélküli státus önmagában rontja az elhelyezkedési esélyeket, részben pszichikai hatása, részben a korábban meglevő emberi tőke elavulása miatt – ezt időtartam-függésnek (duration dependence) nevezik. Vegyük észre, hogy ez a szelekciós torzítás pont ellentétes az előzővel, eredője ezért nem meghatározható irányba torzíthatja a program hatásának a becslését a munkanélküliség időtartamára. A kezelt csoport munkanélküliségi időtartamát ezért nem szerencsés közvetlenül összehasonlítani a kontroll csoportból végig munka nélkül maradottakéval.

Alternatívaként merül fel, hogy az összehasonlítás a kontroll csoport azon tagjainak munkanélküli időtartamával történjen, akik időközben lettek munkanélküliek (újonnan munkanélküliek, “fresh spells”). Minthogy a munkanélküliség a kísérletben való részvétel kritériuma volt, ehhez nekik először munkát kellett találniuk, majd abból ismét munkanélkülivé kellett válniuk. Minthogy igen hátrányos helyzetű csoportokról van szó, ilyen történet viszonylag kevés van, az összehasonlítás ezért imprecíz becslésekhez vezethet. Ráadásul a kontroll csoportból azok, akik munkát találtak, valószínűleg jobb képességűek voltak, és bár akik azután elvesztették az állásukat ez utóbbiak közül átlagosan valószínűleg a rosszabb képességűek, még így is jobbak, mint azok, akik egyáltalán nem találtak munkát. Az kontrollok között az új munkanélküli történetek egyénei tehát valószínűleg átlagosan jobb képességűek. Az egyszerű összehasonlítás itt sem szerencsés, bár legalább a szelekciós torzítás iránya egyértelmű.

Ham és LaLonde kimutatják, hogy a végig munka nélküli kontrollok mérhető ismérveikben nem nagyon térnek el a kezelt csoportból munka nélkülivé válóktól (ez konzisztens a két különböző irányú szelekcióval), az újonnan munkanélküli kontrollok viszont iskolázottabbak, és átlagosan többet dolgoztak korábban.

Az időtartamok vizsgálatánál (duration analysis) hagyományosan alkalmazott modellek a következő két feltevés valamelyikére (vagy mindkettőre) épülnek: vagy nincs időtartam-függés (duration dependence), vagy van, de akkor a nem megfigyelt heterogenitásnak függetlennek kell lennie a megfigyelhető ismérvektől. Ebben az esetben egyik feltevés sem szerencsés, hiszen nem zárhatjuk ki az időtartam függést, és az előbbiekben éppen amellett érveltünk, hogy a nem megfigyelt heterogenitás összefügg a programban való részvétellel, tehát a legfontosabb mérhető változónkkal. Ham és LaLonde ezért innovatív megközelítéssel élt: több modellt becsültek meg, amelyek különböző, nagyon egyszerű feltevéseket tettek a nem megfigyelt heterogenitás eloszlására. A becsült modellek így is rendkívül bonyolultak lettek, ezért itt nem tárgyaljuk őket részletesen.

Az eredmények azt mutatják, hogy a programnak szignifikáns pozitív hatása volt a munkahelyen töltött időre: azt átlagosan körülbelül 11 hónappal hosszabbította meg. A szelekciós problémákat kezelő modellek eredménye viszont azt mutatja, hogy a munkanélküliség időtartamára a program nem volt kimutatható hatással. Az eredmények tehát az (1) magyarázatot részesítik előnyben: az NSW program azáltal növelte a foglalkoztatást, hogy olyan képességekkel ruházta fel a résztvevőit, amelyek miatt azok jobban helyt tudtak állni a munkahelyükön.

6. A KÖZVETETT HATÁSOK MÉRÉSE

Mindeddig azt vizsgáltuk, hogy hatnak-e (és ha igen hogyan hatnak) a vizsgált programok az azokban résztvevőkre. A hatásvizsgálat módszertana ma leginkább csak e közvetlen hatással foglalkozik: az egyik első feltételezés szinte minden módszertani írásban az, hogy a programnak csak az abban részt vevőkre van hatása (lásd például Wooldridge, 2002, 18. fej.).

Egy társadalompolitikai program össztársadalmi hasznai és költségei azonban természetesen a nem résztvevőkre gyakorolt esetleges hatásokat is tartalmazzák. Ha például egy program következtében a résztvevők foglalkoztatási esélyei megnőnek, társadalmi szempontból nyilvánvalóan nem mindegy, hogy a magasabb foglalkoztatás meglévő munkahelyeken való al-

kalmazás eredménye addig ott dolgozók helyett, vagy pedig új munkahelyek teremtésének az eredménye.

A nem résztvevőkre gyakorolt, tehát közvetett hatásokat is magában foglaló, hosszú távú összhatást az irodalomban gyakran „általános egyensúlyi" hatásnak nevezik. Az elnevezés azt takarja, hogy az összhatásban elvileg benne van minden olyan hatás is, amely nemcsak az adott a munkaerőpiacon jelenik meg (akár közvetett akár közvetlen), hanem más munkaerőpiacokon, illetve egészen más piacokon is (pl. termék- és tőkepiacok). Az "általános" elnevezés ambiciózussága ellenére azért a közvetett hatásokan mérni próbáló – egyébként sem túl gyakori – vizsgálatok megelégednek az adott munkaerőpiacon gyakorolt összhatások mérésével. Ezekben az esetekben valószínűleg helyesebb lenne a parciális egyensúly kifejezést használni.

Az "egyensúly" egyfajta hosszú távú hatásra utal. Elképzelhető ugyanis, hogy a program hatására esetleg megnövő és (vagy minőségben eltolódó) munkakínálat felszívásához idő kell, vagyis az egyensúlynak be kell állnia. Általános egyensúlyban különösen fontos ez a kérdés: például más munkaerőpiacokról idővel átcsoportosított tőke hosszú távon az adott munkaerőpiacon megnövelheti a teljes foglalkoztatást, másutt azonban csökkentheti azt. Nem minden közvetett hatás-elemzés egyensúlyi azonban. Ilyen például a programnak a nem résztvevők foglalkoztatására gyakorolt hatását vizsgáló Davidson és Woodbury (1993) tanulmány. Az ő elemzésük csak munkaerőpiaci hatásokkal foglalkozik, és azon belül is feltételezi, hogy bármilyen hatása van is a programnak, annak nincsen hatása a bérekre és árakra. Szigorúan véve az ő megközelítésük tehát nem egyensúlyi, hiszen nem engedik meg, hogy a program esetleges pozitív közvetlen hatására (például valamilyen kínálati görbe jobbra tolódásával) a bérek csökkenjenek, ezáltal növelve a teljes foglalkoztatást (nem teljesen rugalmatlan munkakereslet esetén). Ezt a problémát szem előtt tartva a nem résztvevőkre gyakorolt az általuk vizsgált negatív hatás egyfajta maximális negatív hatásként értelmezhető.

A közvetlen hatást úgy definiáltuk, mint a program hatását azokra, akik részt vettek benne:

$$\mathrm{ATE}_1 = \mathrm{E}(\Delta_i \mid D_i = 1) = \mathrm{E}(Y_{1i} \mid D_i = 1) - \mathrm{E}(Y_{0i} \mid D_i = 1).$$

A teljes (egyensúlyi) munkaerőpiaci hatás ennél többet tartalmaz. Jelöljük T=1-gyel azt a világállapotot, amelyben fut a program, és T=0-val azt, amelyben nem. A teljes hatás ekkor egyszerűen az össztársadalmi szinten várható eredmény különbsége két világállapot között:

$$E(Y_i|T=1) - E(Y_i|T=0) = E(Y_i|D_i=1,T=1)\Pr(D_i=1) - E(Y_i|D_i=1,T=0)\Pr(D_i=1) + E(Y_i|D_i=0,T=1)\Pr(D_i=0) - E(Y_i|D_i=0,T=0)\Pr(D_i=0),$$

ahol D_i=1 azokat az egyéneket jelölik, akik részt vesznek a programban ha az fut, D_i=0 pedig azokat, akik akkor sem. A teljes hatás tehát felbontható a résztvevőkre gyakorolt hatásra (ez nem más, mint az ATE_1), és a nem résztvevőkre gyakorolt hatásra. Általános egyensúlyban a résztvevőkre is rengeteg minden hathat: a program elvileg más piacokon (pl. tőkepiac) is éreztetheti a hatását. A teljes hatás további specifikálása e hatásmechanizmusok pontos körülhatárolását és modellezését igényelné, ami nem célunk ebben az írásban, és az aktív foglalkoztatáspolitikai eszközök hatásának vizsgálattal foglalkozó irodalomban is igen ritkán fordul elő.[21]

6.1 Kísérlet egy kötelező munkakeresés-segítő program teljes hatásának vizsgálatára: Blundell et al. (2002)

Az Egyesült Királyságban az 1990-es évek végén New Deal for Young People néven átfogó programot indítottak el a fiatal hosszú távú munkanélküliek foglalkoztatásának elősegítése érdekében. Blundell, Costa Dias, Meghir és Van Reenen (2002) tanulmánya a program első részének, a munkakeresést segítő "Gateway" programnak a hatását vizsgálja. A program kötelező minden 18-24 év közötti, fél éve munkanélküli segélyben részesülő fiatal számára (a nem részvétel a segély megvonását eredményezi). Mindenkihez rendelnek egy személyes tanácsadót, aki segít és ösztönöz a munkakeresésben, és kisebb tréningeket is tartanak.

A tanulmány eredményváltozóként az elhelyezkedési esélyt vizsgálja a program kezdetétől számított négy hónapon belül. A program résztvevőit rendes állások mellet támogatott állásba is kerülhettek a New Deal program egy más alprogramja keretében. Ebben a munkáltatók tipikusan az összehasonlítható keresetek mintegy 40 százalékát kitevő támogatást kapnak az alkalmazott után. A résztvevőkre gyakorolt (tehát közvetlen) hatáson túl a ta-

[21] Valódi általános egyensúlyi hatáselemzésre a ritka példák egyike Heckman, Lochner és Taber (1998) tanulmánya. A szerzők egy részletesen kidolgozott együtt élő generációs (overlapping generations) általános egyensúlyi modellt építenek fel, amelyben explicit módon modellezik a fizikai tőke és emberi tőke felhalmozást (pl. iskoláztatási döntés), heterogenitást megengedve. A szerzők az Egyesült Államok munkaerőpiacát szimuláló modellt olyan kérdések elemzésére is használják, mint a technológiai fejlődés, a bevándorlás, vagy a különböző méretű kohorszok belépésének a hatása (baby boom). A modell társadalompolitikai programok hatásának vizsgálatára is alkalmas: a szerzők a tandíjak, illetve a tandíj-kedvezmények változásának a hatását elemzik az iskoláztatásra és a kereseti egyenlőtlenségekre, mindenféle indirekt hatást is számításba véve.

nulmány célja az azonos munkaerőpiacon jelen levő nem résztvevők elhelyezkedési esélyeire gyakorolt (tehát közvetett) hatásának a vizsgálata.

A közvetett hatás egyrészt abból fakadhat, hogy a munkakeresésben nyújtott segítség miatt a résztvevők hamarabb találnak munkát. Másrészt a munkahelyi támogatások effektíve csökkentik a munkáltató számára a munkaerő-költséget, így növelve az irántuk támasztott keresletet – potenciálisan más munkavállalók rovására. A tanulmány megfogalmazásában e közvetett hatások "helyettesítési hatások" lehetnek, vagyis abból fakadhatnak, hogy a munkáltatók a célcsoport tagjait mások helyettesítésére használhatja. Természetesen ez a helyettesítési hatás egészen mást takar, mint a tervezett kísérleti hatásvizsgálatokban használt fogalom, ahol a programot levezénylők kompenzáló magatartását értik alatta (az itt használt fogalom közelebb áll a közgazdaságtani terminológiához). A közvetlen és az itt vizsgált közvetett hatások együttesen az adott (helyi) munkaerőpiacra értelmezett egyensúlyi hatást jelentenek.

A kötelező jelleg miatt a hatásvizsgálathoz lehetetlen tökéletes nem kísérleti kontroll csoportot találni. A tanulmány identifikációs stratégiája egy természetes kísérlet megközelítésen alapul. A program 1998 áprilisában indult el teljes egészében, de néhány helyen próba jelleggel három hónappal korábban elindították. A próba-programban részt vevő területeken élők eredményeit így össze lehet hasonlítani a többi területen élő hasonló egyének eredményeivel, legalábbis arra a néhány hónapra, amíg a többi területen nem indult be a program. A közvetlen hatás identifikációjának feltétele az, hogy a kontroll csoportba tartozás (megfelelő változókra kondícionálva) nem függ össze a programban résztvevők várható eredményével. A próba-programban érintett és nem érintett területek összehasonlításánál ez a klasszikus természetes kísérlet feltevésnek felel meg.[22]

A tanulmány a közvetlen és közvetett munkaerőpiaci hatásokat egyszerűen úgy méri, hogy a próba-programban érintett és nem érintett területeken nem csak a program kritériumainak megfelelő egyének eredményeit vizsgálta (a kötelező jelleg miatt az érintett területeken ezek mind résztvevők voltak), hanem a kritériumoknak nem megfelelő egyének eredményeit is (pl. az ép-

[22] A tanulmány egy másik megközelítést is használt, amelyben a kontroll csoport a 24 évesnél valamivel idősebb (25-30), ezért nem részt vevő, de a részvételi kritériumoknak egyébként megfelelő egyének alkotják. A közvetlen hatásnál az összehasonlíthatóság feltétele az, hogy a 19-24 és a 25-30 éves hosszú távú munkanélküliek (megfigyelhető ismérveik kontrollálása után) nem különböznek szisztematikusan a program várható hatásában. A közvetett hatás felfelé torzítaná ezt a különbséget, amennyiben a kritériumoknak megfelelő 19-24 évesek foglalkoztatása úgy nő, hogy a hozzájuk hasonló 25-20 évesek foglalkoztatása csökken.

pen hogy idősebbekét). Megfelelő feltevések mellett ezek az összehasonlítások identifikálják az érintett területeken a nem részt vevőkre gyakorolt hatásokat. Ebben a megközelítésben a teljes hatások területeken belüli hatásokat jelentenek. A vizsgálat így szükségképpen eltekint azoktól a lehetséges mechanizmusoktól, amelyek a többi területet is érintik (pl. a munkaerő vagy a tőke mobilitása).

Az elemzéshez Blundell et al. a brit munkanélküli regiszter adminisztratív adatait használták 1982-től 1999-ig. Többféle módon kiválasztott kontroll területeket is használtak: az összes nem érintett terület mellett egy olyan almintát is, amelynek tagjai mérhető ismérveikben jobban hasonlítottak az érintett területekre (azok az átlagosnál depressziósabb övezetekből kerültek ki). Ökonometriai modellek[23] mellet propensity score alapú párosítási modelleket[24] is becsültek, az eredmények robusztusságát ellenőrizendő.

Az eredmények igen nagy közvetlen hatást mutatnak: a 19-24 éves férfiak elhelyezkedési esélyeit a program körülbelül 10 százalékponttal növelte (az átlagos résztvevő esélyeinek 24-ről 34 százalékra való növekedését jelenti). E 10 százalékpontnyi növekményből a becslések szerint 5-6 százalékpontnyi talált munkát a támogatott foglalkoztatási program keretében, így a maradék 4-5 százalékpontnyi növekedés tudható be tisztán a munkakeresést segítő Gateway programnak. A nem résztvevő csoportokra gyakorolt közvetett hatás a becslések szerint elhanyagolható. Az eredmények specifikációtól függően mutatnak eltéréseket, de a közvetlen hatást elérő negatív közvetett hatást nem lehet kimutatni. Ez vagy azzal magyarázható, hogy a résztvevőket az esetek többségében újonnan kreált munkahelyekre vették fel, vagy azzal, hogy ha ki is szorítottak másokat, az ő foglalkoztatásuk annyival leszorította a hasonló képzettségűek egyensúlyi bérét, amely mellett a többieket is érdemes volt foglalkoztatni a program nélküli világállapotnak megfelelő szinten. A kétféle közvetett hatás ilymódon nagyjából kiegyenlíthette egymást. A szerzők megfelelő adatok hiányában nem vizsgálják a keresetek alakulását, ami közvetlen evidencia lehetne ez utóbbi hatás jelenlétére.

---

[23] A következő ökonometriai modellt becsülték:

$$Y_{ijt} = \theta_j + \delta_t + \gamma' X_{ijt} + \alpha D_{jt} + u_{ijt} .$$

Az i index jelöli az egyént, a j azt, hogy melyik csoportban van (kezelt vagy kontroll), t az idő. Y a bináris eredményváltozó (dolgozik vagy munkanélküli), θ a csoport-specifikus konstans, X az egyéni (időben fix vagy változó) jellemzők vektora, D pedig a kezelést jelző bináris változó. A modellt ebben a lineáris formában, és logit specifikációban is megbecsülték.

[24] Kettős propensity score alapú párosítást végeztek: külön a reform beindítása előtti, és külön a reform alatti időszakra párosították a reform során kezelésbe került és nem kerül területeket, illetve korcsoportokat.

7. KONKLÚZIÓK

A tanulmányban az aktív munkaerőpiaci eszközök hatásvizsgálatának módszertani problémáit és azok megoldási lehetőségeit vizsgáltuk. Elsősorban a közvetlen hasznok mérésének problémáival foglalkoztunk, vagyis azzal a kérdéssel, hogy az adott program milyen és mekkora hasznot hozott a benne részt vevők számára.

A közvetlen hatások arra a kérdésre keresik a választ, hogy milyen előnyökkel járt a részt vevők számára a program ahhoz képest, mintha nem vettek volna részt benne. Ez direkt módon természetesen nem mérhető, hiszen egy egyén vagy részt vett a programban, vagy nem. Ha a gondolatkísérlet nem is, az átlagos hatás mérésére alkalmas kísérletek elvileg elvégezhetők a részvétel véletlenszerű hozzárendelésével. A tervezett kísérletek azonban alapvető morális és politikai problémákat vetnek fel, és általában kis méretben folytathatók csak le. Különböző területi egységek véletlenszerű kiválasztása kísérleti programokra elvileg jobban kezeli ezeket a kérdéseket, ám praktikusan nem old meg mindent. Ugyanez igaz a természetes kísérletekre is, ahol a kiválasztás nem véletlenszerű, ám azzal a feltevéssel élünk, hogy (megfigyelhető ismérvekre kontrollálva) nem függ össze a várható hatásokkal. Bemutattuk, hogy még tervezett kísérletek esetén is további problémákkal jár a hatások mögötti mechanizmusok mérése.

Nem kísérleti szituációkban alapvető kérdés a kontroll csoport kiválasztása és az, hogy milyen gazdag (és a csoportok között mennyire összehasonlítható) a megfigyelhető változók halmaza. Az összehasonlíthatóságnak több olyan kritériumát is számon szokták kérni, amelyek valójában nem szükségesek, ám elégségesek a hihető eredményekhez (azonos adatfelvételből származó résztvevő és kontroll megfigyelések, azonos munkaerőpiacok, stb.). Az összehasonlíthatóság nyilvánvalóan szükséges feltétele a hatások identifikációjának, azonban esetről esetre változó feltételeket szab, és ezeket hüvelykujj-szabályok helyett mindig külön érdemes megvizsgálni.

Az alkalmazható mérési módszerek azonos kontroll csoport és megfigyelhető változók mellett elvileg azonos hatásokat identifikálnak, ha megfelelő körültekintéssel alkalmazzuk őket. A vizsgált módszerek között az az igazán lényeges különbség, hogy a párosításon alapuló ("matching") modellek direkt módon figyelembe vesznek sok olyan feltételt, amikre a nem kísérleti modellek másik típusában, az ökonometriai modellek esetén külön oda kell figyelni.

A közvetett, tehát a programban nem részt vevőkre gyakorolt hatások mérése újabb problémákat vet fel. Egy sikeres program növelheti a résztvevők

elhelyezkedési esélyeit úgy, hogy azok olyan munkahelyekre helyezkednek el, amelyeket a program hiányában mások töltöttek be. Ez a negatív hatás társadalmi szempontból csökkentheti, extrém esetben teljesen kiegyenlítheti a közvetlen pozitív hatást. Az adott munkaerőpiacon kifejtett teljes ("egyensúlyi") hatásban emellett szerepet játszhat az is, hogy az esetleg megnövelt foglalkoztatás (munkakínálat) csökkentheti a béreket – ez ismét a foglalkoztatás növelésének irányába hathat, de alacsonyabb keresetek mellett. A közvetett hatások tipikusan akkor mérhetők, ha a program egyes munkaerőpiacokon bevezetésre került, míg másokon nem. A munkaerőpiacok közötti ("általános egyensúlyi") hatásokat leszámítva ekkor mérhető az érintett munkaerőpiacokon kifejtett teljes (közvetlen és közvetett) hatás.

Összességében elmondható, hogy a társadalompolitikai programok, és így az aktív munkaerőpiaci eszközök hatásának mérése nem egyszerű feladat, és nem minden esetben végezhető el hitelesen. Szerencsés esetben, megfelelő adatok mellett azonban elegendő körültekintéssel általában azonosíthatók a közvetlen hatások és az azok mögött meghúzodó mechanizmusok, és ritkábban a közvetett hatások egy része is.

IRODALOM

Angrist, Joshua D., Guido W. Imbens és Donald B. Rubin (1996): "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association*, 91(434): 444–455.

Ashenfelter, Orley (1978): "Estimating the effect of training programs on earnings," *Review of Economics and Statistics*, 6(1): 47–57.

Ashenfelter, Orley és David Card (1985): "Using the longitudinal structure of earnings to estimate the effect of training programs," *Review of Economics and Statistics*, 67(4): 648–660.

Blundell, Richard és Monica Costa Dias (2002): "Alternative approaches to evaluation in empirical microeconomics," *cemmap Working Paper* CWP10/02.

Blundell, Richard, Monica Costa Dias, Costas Meghir, és John Van Reenen (2002): "Evaluating the employment impact of a mandatory job search program," *IFS Working Paper* WP10/02.

Davidson, C. és S. Woodbury (1993): "The displacement effect of reemployment bonus programs," *Journal of Labor Economics*, 10(4): 380–388.

Dehejia, Rajeev V. (2003): "Practival propensity score matching: A reply to Todd and Smith," *Columbia University Working Paper.*

Dehejia, Rajeev V. és Sadek Wahba (1998): "Causal differences in non-experimental studies: Re-evaluating the evaluation of training programs," *NBER Working Paper 6586.*

Dehejia, Rajeev V. és Sadek Wahba (2002): "Propensity score matching methods for non-experimental causal studies," *Review of Economics and Statistics*, 84(1): 151–161.

Hahn, Jinyong (1998): "On the role of the propensity score in efficent semiparametric estimation of average treatment effects," *Econometrica*, 66: 315–331.

Ham, John C. és Robert J. LaLonde (1996): "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training." *Econometrica*, 64(1): 175–205.

Heckman, James J. (1996): "Notes on Angrist, Imbens, and Rubin, Identification of causal effects using instrumental variables," *Journal of the American Statistical Association*, 91(434): 460–463.

Heckman, James J., H. Ichimura, Jeffrey A. Smith és P. Todd (1997): "Characterising selection bias esing experimental data," *Econometrica*, 66(5): 1017–1098.

Heckman, James J., H. Ichimura, és P. Todd (1997): "Matching as an econometric evaluation estimator," *Review of Economic Studies*, 64: 605–654.

Heckman, James J., Robert J. Lalonde és Jeffrey A. Smith (1999): "The economics and econometrics of active labor market programs," in: A. Ashenfelter and D. Card, eds, *Handbook of Labor Economics, Vol. 3A*, Elsevier, 1865–2097.

Heckman, James J., Lance Lochner és Christopher Taber (1998): "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogenous agents," *Review of Economic Dynamics,* 1(1):1–58.

Heckman, James J., Jefrrey A. Smith, és Nancy Clemens (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *Review of Economic Studies*, 64(4): 487–535.

Hirano, K., G. W. Imbens és G. Ridder (2000), "Efficient estimation of average treatment effects using the estimated propensity score," UCLA Working Paper

Imbens, Guido W. és Joshua D. Angrist (1994), "Identification and estimation of local average treatment effects," *Econometrica*, 62(2): 467–475.

LaLonde, Robert J. (1986): "Evaluating the econometric evaluations of training programs with experimental data." *American Economic Review*, 76(4): 604–620.

Rosenbaum, P. és D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70: 41–55.

Rosenbaum, P. és D. B. Rubin (1984): "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Association*, 79: 516–524.

Smith, Jeffrey A. és P. Todd (2002): "Does matching overcome LaLonde's critique of nonexperimental estimates?" *nem publikált műhelytanulmány*.

Willis, Robert J. és Sherwin Rosen (1979): "Education and self-selection," *Journal of Political Economy*, 87(5/2), S7-S36.

Wooldridge, Jeffrey M. (2002): *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge MA.