

BALÁZSI LÁSZLÓ–DIVÉNYI JÁNOS KÁROLY–  
KÉZDI GÁBOR–MÁTYÁS LÁSZLÓ

## A közgazdasági adatforradalom és a panelökonometria

Áttekintő tanulmányunk célja, hogy – a modern panelökonometria történetén és fejlődésén keresztül – bemutassa, milyen általános és módszertani következményekkel járhat a jelenleg zajló közgazdasági adatforradalom. Látni fogjuk, hogy a fejlődés valójában alulról építkezik: a paneltípusú adatok megjelenése különböző problémákat hoz felszínre, amelyekre elméleti és módszertani újításokat adó válaszok születnek, s az újítások aztán visszahatnak újabb és egyre bonyolultabb adatbázisok gyűjtésére. Ez a visszacsatolási mechanizmus a legfontosabb mozgatórugója a jelenleg zajló, úgynevezett *big data* adatforradalomnak, amelynek eredményeképpen mélyebb s egyszerre átfogóbb és részletesebb képet kaphatunk a gazdaság egészének működéséről.

Journal of Economic Literature (JEL) kód: A11, B23, N01, O30.

Közhely napjainkban, hogy a közgazdaságtan forradalmi változások küszöbén áll, ami elsősorban a vizsgálatokhoz rendelkezésre álló adatbázisok robbanásszerű növekedésének és elterjedésének köszönhető. Ez a hihetetlen adatmennyiség olyan új összefüggések feltárását és kvantifikálását teszi lehetővé, amire eddig gondolni sem mertünk. Az adatforradalom ugyanakkor számos új problémával is járhat. Egyrészt nagyobb mennyiségű adatból sokkal nehezebb megtalálni a relevánst és kiszűrni az esetlegest, másrészt a mennyiségi növekedés mellett az adatok szerkezete is egyre bonyolultabbá válik, ami egyre bonyolultabb elemzési módszereket kíván. Ráadásul az újszerű adatok jelentős része „zajos”, vagyis hemzseg a mérési hibáktól, és rengeteg bennük a hiányzó adat. A módszertani fejlődés mára olyan szintet ért el, hogy gyakorlatilag csak az erre specializált szakértők tudják követni a legújabb modelleket és trendeket. Sok esetben az alapokat is újra kell gondolni, hiszen gyakran a jól bejáratott módszereket nem lehet a gyakorlatban alkalmazni a hatalmas méretű, összetett szerkezetű és sok irreleváns információt (is) tartalmazó adatokon.

---

Balázi László, Közép-európai Egyetem, Közgazdaságtan Tanszék (e-mail: Balazsi\_Laszlo@ceu-budapest.edu).

Divényi János Károly, Közép-európai Egyetem, Közgazdaságtan Tanszék (e-mail: divenyi\_janos@ceu-budapest.edu).

Kézdí Gábor, Közép-európai Egyetem, Közgazdaságtan Tanszék (e-mail: kezdig@ceu.hu).

Mátyás László, Közép-európai Egyetem, Közgazdaságtan Tanszék (e-mail: matyas@ceu.hu)

A napjainkban keletkező megannyi közgazdasági és üzleti adatbázis sokszor paneladatok formájában ölt testet. Egyrészt mert gyakran természetükből adódóan ilyen jellegűek, másrészt mert számtalanszor a tiszta idősorokat vagy tiszta keresztmetszeti adatokat is célszerű ilyen struktúrájúvá alakítani.

A jelen írás célja a panelökonometria szemüvegén, példáján és történetén keresztül bemutatni, hogy az új adatforradalom milyen módszertani következményekkel járhat, illetve milyen alapokra épülhet a módszertan továbbfejlődése. A forradalmakat nehéz menetközben megérteni, de a főbb trendeket, reméljük, jól ismerjük fel.

Az olvasók többsége bizonyára már találkozott a paneladatbázis vagy az erre épülő tudományág, a panelökonometria fogalmával. Az 1960-as évektől gyakorlatilag áradatszerűen megjelenő cikkekén kívül több jeles összefoglaló könyv is segíti a téma megértését és követését. Kevés rendszerezett információnk van azonban arról, hogy mi is történt a panelökonometria elmúlt körülbelül egy évtizedében, illetve milyen új irányban, mely területek felé halad jelenleg a tudományág.

Azok kedvéért, akik kevesebb háttérismerettel rendelkeznek ebben a témakörben, elevenítsük fel, mit is jelent a fogalom! Általános értelemben minden olyan adatbázis paneladatbázis, amelyben a megfigyelések többdimenziós struktúrába rendezhetők. A hagyományosan leggyakrabban elemzett paneladatbázisban egyedek egy adott halmazát figyeljük meg rögzített időpontokban egy adott időn át. Megfigyeléseink tehát egy adott egyed valamilyen tulajdonságaiból és ezek változásaiból állnak. Az ilyen szerkezetű adatbázisokra épülő és – ami igazán fontos – az ebből a szerkezetből adódó lehetőségek kihasználásával foglalkozó ökonometriai módszereket nevezzük panelökonometriának. Tekintsünk egy egyszerű példát! Szeretnénk számszerűsíteni a GDP importra kifejtett hatását, vagyis azt, hogy a GDP növekedése (csökkenése) milyen mértékben növeli (csökkenti) az importot. Egy ilyen kutatás alapjául szolgáló paneladatbázisban az egyedek országok, amelyek importját, GDP-jét stb. figyeljük meg adott időpontokban (mondjuk negyedévente). Az így képzett adatbázisra írjuk fel a modellünket, amelynek segítségével megkapjuk a vizsgált hatás becslését. A legfontosabb kérdés az egyes országok közötti heterogenitás. Egyrészt a különböző GDP-vel rendelkező országok sok más tényezőben különbözhetnek egymástól, és ezek önmagukban is oksági kapcsolatban állhatnak az importtal. Másrészt a GDP és az import közötti kapcsolat is különbözhet országok, országcsoportok között (kis országok, nagy országok stb.). A panelmodellek egyik legfontosabb előnye, hogy megfelelő körülmények – és helyesen felírt modell – esetén az ilyen jellegű heterogenitást képes kezelni.

Egy ilyen adatbázis tipikusan változónként legfeljebb néhány ezer megfigyelést tartalmaz. Ha azonban egyes országok helyett az országpárok közötti importot vagy exportot (kereskedelmet) tekintjük egy egyednek, és ennek összefüggéseit akarjuk vizsgálni, akkor világméretű adatbázis esetén egymilliót megközelítő megfigyelésünk is lehet. Az utóbbi években már ágazati bontásban is készülnek export–import adatbázisok, ami tovább növeli a megfigyelések számát. Az esetszámok növekedése mellett az adatszerkezet is egyre bonyolultabbá válik: egyedek több időpontbeli megfigyelése helyett ágazatok országpárok szerinti, több időpontbeli megfigyeléséről van szó. Ez kiváló példa az adatforradalom folyamatára.

Visszalépve egy kicsit: felmerülhet a kérdés, miért érdekes egy panel, mennyivel nyújt többet, mint mondjuk egy hasonló méretű keresztmetszeti adatbázis. Tegyük fel, hogy szeretnénk háztartások fogyasztási szokásait elemezni egy adott időpontban. Ehhez felhasználhatunk egy keresztmetszeti adatbázist, amelyben különböző háztartásokról különböző adatok állnak rendelkezésre azonos időpontban. Ha egy adott háztartás megtakarítási szokásait szeretnénk figyelemmel kísérni hosszabb időn keresztül, idősoros adatbázisról van szó. Ez a két adatbázis azonban csak nagyon korlátozottan alkalmas oksági összefüggések feltárására: keresztmetszet esetében a háztartások közötti nem megfigyelt heterogenitás és a fogyasztás összefüggései okoznak gondot, idősorban pedig a környezet nem megfigyelt változásai. Ha több háztartást figyelünk meg különböző időpontokban, paneladatbázisunk van, ami közelebb vihet az oksági összefüggések megértéséhez. Ki tudjuk szűrni az egyes egyénekre jellemző, nem megfigyelhető ismérveket („kontrollálhatunk rájuk”), ha azok időben állandók (egyéni hatások), illetve az aggregált trendeket is, anélkül hogy specifikálnunk kellene azok formáját (időhatások), továbbá számos különböző típusú kölcsönhatást is. Nem meglepő tehát, hogy a paneladatbázisok összehasonlíthatatlan népszerűsége tettek szert néhány évtized leforgása alatt, és módszertanuk, ökonometriájuk a közgazdaságtan egyik legdinamikusabban fejlődő ágává vált.

A következőkben először röviden összefoglaljuk a panelökonometria történelmét és jelentősebb mérföldköveit, hogy megértsük a jelenleg zajló robbanás gyökereit. Utána a legfrissebb vívmányokat tekintjük át, kitérve a módszertanra, a legnépszerűbb területekre és a legfontosabb adatbázisokra. Végül megpróbáljuk felvázolni a panelökonometriai szakirodalom esetleges közeli jövőképét, és igyekszünk eligazítani a nagyobb adatbázisok között.

## A panelökonometria története

A panelökonometria gyökereit *Nerlove* [2005] szerint *Airy* [1861] könyvében találhatjuk. Az angol matematikus és csillagász a csillagászati megfigyelésekkel kapcsolatban jegyezte meg, hogy a különböző napokon elvégzett, de azonos célra irányuló mérések (például a Jupiter átmérője) a szükséges korrekciók után is eltérnek egymástól. Ebből az eltérésből azt a következtetést vonta le, hogy minden egyes napon olyan speciális okoknak kell közreműködniük, amelyek csak az adott napra jellemzők. Ezek a speciális okok felelősek a mérések máshogy nem indokolható eltéréseért, amelyeket ő konstans hibának nevezett. Az a felismerés, hogy az egyes megfigyelésekben található esetlegességekben (más szóval zajban vagy hibában) megfigyelhető egyfajta szabályszerűség (például a fenti esetben az ugyanazon napon végzett mérések összefüggnek, szisztematikusan különbözhetnek más napokon végzett mérésektől), a paneladatok használatának egyik leglényegesebb eleme. *Airy* konstans hibáját mai terminológiával élve véletlennap-hatásnak hívnánk.

Jó 50 évvel ezt követően az eredetileg biológusként induló Ronald A. Fisher megfigyelésekben rejlő esetlegességek szabályszerűségeinek strukturált feltárására, azaz a hiba varianciájának modellszerű felbontására irányult két, kvantitatív genetikával foglalkozó

tanulmányában írta le az azóta is népszerű varianciaanalízist (Fisher [1918a], [1918b]), amelyet szokás ANOVA-módszernek nevezni (ANalysis Of VAriance). Néhány évvel később megjelenő korszakalkotó könyvében (Fisher [1925]) a panelökonometria területén máig alapvető fix hatás és véletlen hatás fogalmai is megjelennek, bár az ezek közötti különbségtételben Nerlove [2005] szerint Fisher sosem volt igazán világos.

A fix és véletlen hatás fogalmának tiszta elkülönítése Eisenhart [1947] a Biometrics című folyóiratban megjelent tanulmányához köthető. A matematikus az adatelemzési problémáknak két osztályát különböztette meg: az „univerzum” összes objektumának egyes részcsoportjaira jellemző átlagok közötti fix (állandó) kapcsolatok feltárását és becslését, illetve egy összetett populációban megfigyelhető véletlen változatosság összetevőinek feltárását és becslését. A két megközelítés közötti választást abból kiindulva kell meghozni, hogy a fix kapcsolatok paramétereire vagy a véletlen variabilitás összetevőire vagyunk-e kíváncsiak. Míg az (abban az időben főként a mezőgazdaságban jellemző) kísérletek az előbbi, a kísérlettel nem megvalósítható (például csillagászati) megfigyelések elemzése inkább az utóbbi modell alkalmazását kívánja meg. Az Eisenhart által megkülönböztetett két problémaosztályra Scheffé [1956] tanulmánya alkalmazza először a fixhatás- és a véletlenhatás-modell kifejezést.

Az egyes hatások (akár fix, akár véletlen) elkülönítésére szolgáló panelmodell (azaz az idő- és csoportdimenziót egyaránt tartalmazó modell) alapjait Hildreth [1950] műhelytanulmányában találhatjuk. A háromkomponensű modell két változata jelenik meg: az idő- és a csoportthatást fix hatásként kezelő fixhatás-modell (a harmadik komponens a nyilvánvalóan véletlen hatásként kezelt általános hibatag), illetve a mindhárom hatást véletlenként kezelő véletlenhatás-modell. Hildreth levezeti a szükséges becslőfüggvényeket, ám megállapítja, hogy a véletlenhatás-modell bonyolult – nemlineáris maximum likelihood – becslőfüggvénye (az akkori technológia mellett) nem megoldható, s így a standard ANOVA-módszerrel könnyen becsülhető fixhatás-modell használatát javasolja. A számítási módszerek akkori fejlettségéből fakadó megkülönböztetés az egész időszakot végigkísérte. Rao [1952] és Scheffé [1959] összefoglaló kötetei egyértelműen a fixhatás-modellre koncentrálnak, a véletlenhatás-modell sokkal kisebb hangsúlyt kap.

Fontos közgazdaságtani probléma panelökonometriai megoldására először Irving Hoch 1957-es PhD-disszertációjában találunk példát [a fő eredményeket Hoch az Econometricában publikálta néhány évvel később (Hoch [1958] és [1962])]. Marschak–Andrews [1944] megmutatta, hogy a termelési függvények akkori becslésének alapvető problémája, hogy az input- és az outputszintek megválasztásáról szóló döntés endogén (például különböző szaktudás, az időjárás nem megfigyelt különbsége stb.), ezért a keresztmetszeti adatokon becsült kapcsolat torzított eredményt ad. Irving Hoch Hildreth [1950] ajánlásaira támaszkodva 63 minnesotai farm hat évre kiterjedő adatain vizsgálja a kérdést, s a keresztmetszeti és idősoros adatok összekapcsolásával elkülöníti a farmra és az egyes évekre jellemző fix hatásokat. Ez az újszerű becslési módszer adta Mundlak [1961] híres cikkének alapját is.

Balestra–Nerlove [1966] az Egyesült Államok tagállamainak gázfogyasztását tartalmazó paneladatokon vizsgálta a gáz iránti keresletet. A probléma dinamikus jellegét felismerve, a szerzők modelljükbe magyarázó változóként bevonták a függő változó

késleltetettjét, ám az akkor megvalósítható fixhatás-módszerrel irreális eredményeket kaptak. Így ismerték fel, hogy a függő változó késleltetettjét magyarázó változóként tartalmazó modell standard fixhatás-modellként történő becslése torzított lesz. A probléma kikerülésére kétkomponensű véletlenhatás-modell becslésére tértek át, amit végül a maximum likelihood módszer számítási problémáinak változatlan fennállása miatt a megvalósítható általánosított legkisebb négyzetek (*Feasible Generalized Least Squares, FGLS*) és az instrumentális változók módszerére alapuló eljárással oldottak meg.

Az elkövetkező évek irodalma főként a hasonló jellegű módszertani problémák megoldására összpontosított, a maximum likelihood módszert kiváltó, minél hatásosabb módszerek keresése került a középpontba (lásd például G. S. Maddala korabeli munkáit). Emellett egyre több alkalmazott tanulmány látott napvilágot (például Mazodier 1971-ben és Chamberlain 1975-ben elkészült harvardi PhD-disszertációja). Felismerték, hogy a panelmódszerek alkalmasak lehetnek látens vagy nem megfigyelt változókat tartalmazó modellek becslésére (például az általános képességek figyelembevételére az iskolázottság bérekre gyakorolt hatásának megfigyelésekor – *Chamberlain–Griliches* [1975]). Mind az empirikus, mind a módszertani irodalom fejlődését felgyorsította a paneladatok elérhetőségének nagymértékű javulása. 1968-ban kezdték el a Michigani Egyetem híres jövedelemdinamika-felvételét (*Panel Study of Income Dynamics, PSID*), amelyet az elsők között *Lillard–Willis* [1978] elemzett.

1977-ben Pascal Mazodier Párizsban megszervezte a panelökonometria első konferenciáját, amely *Nerlove* [2005] szerint óriási hatással volt az elkövetkező évek tudományos fejlődésére. Ez a fejlődés jórészt két forrásból táplálkozott: egyrészt folytatódott a paneladatok elérhetőségének javulása, másrészt felismerték, hogy a kvantitatív elemzések elengedhetetlenek a gazdasági fejlődéshez szükséges szakpolitika kidolgozásához, s mivel a probléma természeténél fogva dinamikus, megoldása legfőképpen panelmódszerek alkalmazását igényli.

A panelökonometria gyors fejlődésének köszönhetően hamarosan, 1987-ben, sor került a második konferenciára is (szintén Párizsban), s ettől kezdve nagyjából szabályos rendben, minden második évben megszervezik az eseményt. Erre az időszakra a panelökonometria már önállóan elismert tudományterületként működött. Olyan tanulmányok születtek a témában, mint például a fix- és véletlenhatás-specifikáció közötti választásban segítő Hausman-próba kidolgozása (*Hausman* [1978]), vagy a mikroökonometriai problémákra jobban illő diszkrét eredményváltozós modellek panelalkalmazásai (például *Maddala* [1987]).

Az első konferencia után jó néhány évvel kezdtek megjelenni az úttörő összefoglaló jellegű könyvek (például *Hsiao* [1986], *Baltagi* [1995], *Mátyás–Sevestre* [1995]), jelezve, hogy a panelökonometria önálló tudományterületté vált.

## Az elmúlt évtizedek fejlődése

A panelökonometria születése óta klasszikusan kétdimenziós adatbázisokra támaszkodott. Egyrészt ez volt a legegyszerűbb adatbázistípus, amely egyaránt rendelkezett a keresztmetszeti és idősoros adatbázis tulajdonságaival, másrészt méretét tekintve

is kezelhető volt. Az elmúlt egy-két évtized rohamos technológiai fejlődése azonban magával hozta az extrém méretű adatbázisok elterjedését. A méreten érthetjük a megfigyelések mennyiségét (hiszen elektronikus úton már nemcsak tárolni, de gyűjteni is könnyebb az adatokat), de akár a magasabb dimenziókba való átlépést is. Az igény az effajta komplex adatbázisokra gyakorlatilag együtt született a tudományággal, leg híresebb példája talán az 1968-tól létező *Blue Chip Survey of Professional Forecasters* (lásd *Bauer és szerzőtársai* [2003]). Az ilyen nagy adatbázisok gyakorlati használhatóságát azonban erősen korlátozta a magas számolási igény. Például: míg száz cég tíz éven át tartó (negyedéves) megfigyeléseiből álló kétdimenziós adatbázis 4000 megfigyelést tartalmaz, úgy ennek egy lehetséges háromdimenziós formája (melyben egy adott cég szállít egy másikkal egy adott időpontban) már 400 ezret. Hasonló ugrásokat tapasztalhatunk, ha háztartási szintű paneladatbázisokat egyéni szintűvé bővítünk, vagy évenkénti megfigyeléseinket pontosítva például heti bontásban gyűjtjük adatainkat. Az egyre bonyolultabb szerkezetű adatbázisok új lehetőségeket, de egy szersmind új kihívásokat is jelentenek. Az első kihívás a számításigények növekedése. A mai számítógépek már könnyen kezelnek nagyobb adatbázisokat, de a megfigyelések és dimenziók számát növelve hamar túlléphetjük az elvégezhető műveletek számára vonatkozó korlátot. A panelökonometriai modellezés tehát kritikus hangsúlyt helyez a számításigényre, illetve a műveletek elvégezhetőségére is. Olyan eljárásokra van szükség, amelyekkel a műveletigény lényegesen redukálható, hiszen hiába vannak papíron kivitelezhető módszereink, ha azok a gyakorlatba átültethetetlenek. Ennek jeles példája a skaláris felírás használata költséges mátrixműveletek helyett. Másrészt a bonyolultabb szerkezetű paneladatbázisok a modellezést is új kihívások elé állítják. A hatások és interakciójuk, valamint a paraméterek esetleges heterogenitásának modellezése sokkal komplexebb, és az eredmények értelmezése is gyorsan válik nagyon nehezzé. Harmadrészt a nagyobb adatmennyiségben nehezebb kiszűrni, hogy köztük mi a releváns. Ebben a tekintetben az ökonometriai alkalmazásoknak van hova fejlődniük, de támaszkodhatnak a természettudományokban – például a genetikai biológiában – zajló folyamatokra is.

Az következőkben bemutatjuk a közgazdaságtanban használt legismertebb paneladatbázisokat, a legfontosabb elméleti eredményeket és azokat a területeket, amelyek a legnagyobb sikerrel alkalmazzák a panelökonometriát.

### *Adatbázisok*

Az elérhető adatbázisok tartalma és szerkezete alapvetően meghatározza, milyen módszerek használatára van igazán szükség, illetve lehetőség. A módszertani kutatókat gyakran újszerű adatbázisok megjelenése sarkallja az elméleti háttéranyag bővítésére vagy éppen finomítására. Így alakult ki, hogy fix és véletlen hatásokat használunk modelljeinkben, ezért jutottunk el a dinamikus modellekhez, ezért tudjuk különböző feltételek esetén is torzítatlanul és konzisztensen becsülni a paramétereket. Érdemes tehát először a legismertebb paneladatbázisokat bemutatni, és csak azután rátérni az új módszertanra.

A leggyakrabban használt paneladatbázisok még mindig háztartások és egyének fogyasztási szokásaival, jövedelmével, foglalkoztatottsággal foglalkozó, nemritkán több évtizedre visszanyúló adatbázisok. Alapvetően két típusuk létezik: az egyéni panelek és a háztartáspanelek. Az egyéni panelek egyéneket követnek, ideális esetben életük végéig. Az egyéni panelek egy részét „frissítik”, vagyis időnként újabb egyénekkel töltik fel, ezzel biztosítva a folyamatosságukat. A háztartáspanelek megfigyelési egysége a háztartás, és akkor is követi a háztartásban már megfordult egyéneket (vagy egy részüket), ha már elköltöztek. Jó tervezés mellett és ideális esetben a háztartáspanel az újabb generációk háztartásaival automatikusan „frissül”.

A legrégebbi, ezért leghosszabb háztartáspanel az 1968 óta működő *Panel Study of Income Dynamics* (PSID – lásd Hill [1992]). A PSID több mint 18 ezer egyén (majd idővel azok leszármazottai) jövedelméről, foglalkoztatottságáról, iskolázottságáról, gyermekvállalási szokásairól rendelkezik adatokkal. Az adatbázis korlátlanul elérhető, talán ez is közrejátszik abban, hogy máig több mint 3000 publikáció alapjául szolgált! A főbb kérdések, melyeket megválasztak, természetesen, a jóléttel kapcsolatosak. Csak néhány példa: Milyen tényezők játszanak közre iskolaválasztás során (Evangelos–Peters [1998]). Öregedik-e a társadalom (Smith [1994]). Milyen a jövedelemeloszlás az Egyesült Államokban (Wolff [1999]). Általános összefoglaló tanulmányok szegénységről és jólétről (Gottschalk–Ruggles [1994]). A PSID lehetőséget adott a korábbi (vagy kisebb) adatbázisokon már vizsgált kérdések újraértékelésére és összehasonlítására is.

A PSID sikerei több országot is arra ösztönöztek, hogy elkészítsék a PSID-hez hasonló saját adatbázisukat. Több mint 40 ezer háztartás szociális és gazdasági körülményeiről gyűjt információt a már öt éve létező *Understanding Society*, amely a mára már inaktív, de 18 évig működő *British Household Panel Survey* (BHPS – lásd Wagner és szerzőtársai [2007]) alapjain nyugszik. A ráépülő kutatások foglalkoznak többek között etnikai csoportok asszimilációjával (Nandi–Platt [2013]), munkavállalási szokásokkal, vagy akár olyan speciális kérdésekkel is, mint a háztartások reakciója a házastárs munkahelyének elvesztésére (Gush és szerzőtársai [2013]).

Hasonló méretű az 1984 óta működő *German Socio-Economic Panel* (GSOEP), mely 11 ezer német háztartás tagjairól (legyen az őslakos, bevándorló vagy külföldi) rendelkezik adatokkal. Széles körű kutatásai a keretekkel is foglalkoznak, de sokszor jóval általánosabb témákat is felölelnek, például hogy milyen kapcsolatban áll a munka és az általános boldogság (Coad–Binder [2014]), vagy boldogtalansághoz vezet-e az intézeti nevelés (Lemola–Richter [2014]). Hasonló felmérésekkel találkozhatunk Koreában (*Korean Labor and Income Panel Study*, KLIPS – lásd KLIPS [2007]), Ausztráliában (*The Household, Income and Labour Dynamics in Australia*, HILDA – lásd Wooden és szerzőtársai [2002]), Új-Zélandon (*Survey of Family, Income and Employment*, SoFIE – lásd Carter és szerzőtársai [2010]), vagy éppen Kanadában (*Survey of Labour and Income Dynamics*, SLID – lásd Giles [1999]). A Tárki az 1990-es években Magyarországon is gyűjtött háztartáspanel-adatokat, ez volt a valamivel kisebb méretű *Magyar Háztartáspanel*, MHP – lásd Sik–Tóth [1998] és Spéder [2001].

Az egyéni panelek közül a legismertebbek közé tartozik az amerikai *National Longitudinal Surveys of Youth* (NLSY), amely egy fiatal korosztály iskolai és korai munkapiaci tapasztalatairól gyűjt adatokat. Két ilyen kohorszot követnek: az NLSY79 az 1979-ben

tizenévesekről, NLSY97 az 1997-ben tizenévesekről gyűjt adatokat. Az NLSY79-ben részt vevők gyermekeit újabb panelvizsgálat kezdte el követni. Az NLSY-felvételek teték először lehetővé olyan kérdések elemzését, mint hogy a különböző etnikai csoportokba tartozó munkavállalók között mért bérkülönbség mennyiben magyarázható a tizenéves korban mért készségek különbségével (Neal–Johnson [1998]), a fiatalok letartóztatása mennyiben hat későbbi munkapiaci esélyeikre (Grogger [1995]), vagy hogy az úgynevezett nem kognitív készségek (mint például a pozitív önértékelés) mennyiben határozzák meg a munkapiaci sikerességet (Heckman és szerzőtársai [2006]). Az NLSY-nak is van magyarországi megfelelője: a 2006-ban, szintén a Tárci által gyűjtött Életpálya-felvétel 10 ezer nyolcadik osztályos fiatalot követ évenkénti megkérdezéssel (Hajdu és szerzőtársai [2014]). Az ausztrál munkaerő-felmérés (*Labour Force Survey, LFS* – lásd Trewin [2005]) az 1960-as évektől kezdve gyűjti össze az adott lakosság munkaerő-piaci aktivitását és azok jellemzőit, különös figyelmet szentelve a munkanélküliségi rátának és a munkanélküliség időtartamának. Ráépülő kutatásaik főként a betanulások, tréningek hosszával és azok hatásaival, különböző korú (fiatal vagy éppen nyugdíj közeli) alkalmazottak munkavállalási szokásaival foglalkoznak.

Kutatók körében nagy népszerűségnek örvendenek az idősebb népességet vizsgáló egyéni panelek is. Az úttörő felmérés ezek esetében is az Egyesült Államokból származik, ez az 1990-es évek óta folyó *Health and Retirement Study (HRS)*, amely az 50 évesnél idősebbeket követi két évenkénti megkereséssel, és a standard háztartáspanelkérdések mellett lefed rengeteg, az idősödéssel kapcsolatos információt is, elsősorban a nyugdíjjal, az egészségi állapottal, az egészségügyi ellátás igénybevételével és az időseknek nyújtott, illetve általuk adott pénzbeli és egyéb támogatásokkal kapcsolatban (lásd például Coile–Levine [2006], Adams és szerzőtársai [2003] vagy McGarry [1999]). A HRS életük végéig követi az egyéneket, és a minta hat évente frissül, új 51–56 évesek bevonásával. Az HRS sikerét jelentős részben annak köszönheti, hogy a szakmai irányítását teljes egészében kutatókból álló csoport végzi. A felvételt – a PSID-hoz hasonlóan – a Michigani Egyetem Survey Research Centere végzi, de tartalmi szempontból egy közgazdászokból, epidemiológusokból, szociológusokból és pszichológus kutatókból álló egyetemek közötti csoport irányítja. Ez nemcsak azt eredményezi, hogy a felmérés a kutatási kérdésekhez igazodó tartalommal és módszerekkel gyűjt információkat, hanem azt is, hogy rengeteg benne az innováció is (például a kockázatpreferenciák mérése, lásd Barsky és szerzőtársai [1997]; vagy a várakozások mérése, lásd például McGarry [2003]). Az HRS-nek mind tartalmában, mind irányítási módjában számos követője akadt, köztük a nagy-britanniai ELSA, a koreai KLoSA, a japán JSTAR, a kínai CHARLS, a mexikói MHAS. Az HRS legambiciózusabb követője a 20 országban teljes mértékben harmonizált felmérés: *Survey of Health, Aging and Retirement in Europe (SHARE)*. A SHARE-hez 2011-ben Magyarország is csatlakozott (a Tárci által gyűjtött magyarországi adatokból publikált Divényi–Kézdi [2013]), ám belföldi források hiányában a későbbi részvételünk meghiúsult.

A makroökonómiai adatbázisok terén igen jelentős helyet foglal el a szintén amerikai *Survey of Professional Forecasters*. A már 1968-tól létező negyedéves felméréseken nyugvó adatbázis 32 változón keresztül méri az amerikai gazdaság alakulását, és ad lehetőséget a legprecízebb gazdasági elemzésekre és előrejelzésekre



(*Baghestani* [2011], *Choy és szerzőtársai* [2006]). Az elmúlt néhány évben egyre nagyobb népszerűségnek örvendenek a különböző kereskedelmi és keresztárfolyami adatbázisok is.<sup>1</sup>

Végül, de nem utolsósorban, említést kell tennünk arról, hogy számos olyan standard adatbázisnak van „paneldimenziója”, amit elsősorban keresztmetszeti adat-szolgáltatásra használnak. A munkanélküli-ráta mérésére használt munkaerő-felvételek általában több tízezer háztartásról gyűjtenek adatot havi vagy negyedéves rendszerességgel, és ugyanazt a háztartást többször – Magyarországon hatszor – is megkeresik. Hasonlóképpen, a háztartások fogyasztási szokásait felmérő felmérések – Magyarországon a Háztartási Költségvetési Felvétel – néhány éven keresztül vesznek fel adatokat ugyanattól a háztartástól. Ezek az úgynevezett rotációs panel-adatbázisok az egyénekről nem hosszú idősorokat, hanem csak néhány megfigyelést tartalmaznak – a minta mindig új háztartások bevonásával frissül –, de méretük és nemzetközi elterjedtségük miatt rendkívül népszerűek az elemzők körében (lásd például *Blanchard–Diamond* [1990]).

A „hagyományos”, kérdőíves módszerrel gyűjtött paneladatok mellett egyre több elemzés épül eredetileg nem kutatási céllal gyűjtött, de a kutatók számára megnyitott úgynevezett adminisztratív adatbázisokra. Az adminisztratív adatbázisok általában egy-egy ország teljes populációjáról tartalmaznak adatokat, teljes idősorban. Tipikus példái a munkanélküli-ellátásban részesülők sokaságáról napi szinten információt tartalmazó munkanélküli-regiszterek, vagy az adóhatóság, a társadalombiztosítás, illetve a vámhatóság által minden befizetőről rendszeresen gyűjtött adatok. Különösen értékesek a különböző szinteken gyűjtött információkat összekapcsoló adatbázisok, mint a kapcsolt munkáltatói és munkavállalói adatok, vagy az export- és importszállítmányok adatainak, illetve a banki hitel-, valamint betéti szerződéseknek az összekapcsolása vállalati mérlegadatokkal. A skandináv országokban elemzések céljára összekapcsolható szinte minden egyéni szintű adminisztratív adat a személyi számon keresztül, a születési, iskolai és egészségpénztári adatoktól az adózási, sorkatonai vagy munkanélküli-regiszter adatokig. Ilyen adatokon vizsgálták például a testvérek számának és a születési sorrendnek a hatását az iskolai végzettségre (*Black és szerzőtársai* [2005]), az öröklés és a nevelés szerepét a társadalmi status átörökítésében, vér szerinti, illetve nevelőszülők és gyermekeik vizsgálatával (*Björklund és szerzőtársai* [2007]) vagy a háztartások vagyonának hatását a kockázatosabb befektetési döntésekre (*Calvet és szerzőtársai* [2008]). Az adminisztratív adatbázisok nyilvánvaló előnye a hatalmas méret, szerkezetük azonban sokszor bonyolult, így elemzésük bonyolultabb panelökonometriai technikákat igényel.

Nyilvánvaló előnyeik mellett az adminisztratív adatbázisok általában kevesebb információt tartalmaznak, mint a nagy kérdőíves adatbázisok. A skandináv országokon kívül például az adminisztratív adatok alapján általában nem lehet megmondani, hogy ki kivel alkot egy háztartást, mi az egyének közötti családi kapcsolat stb. Az adminisztratív adatbázisokban nem szerepelnek szubjektív kérdések (amelyek

<sup>1</sup> A legfontosabb adatbázisokról találhat az olvasó egy gyűjteményt a <http://www.paneldataconferences.org> weblapon a *Resources* fül alatt.

lehetővé tennék a vélekedések, preferenciák stb. mérését), és sok viselkedésnek sincs nyoma bennük (politikai viselkedés, időfelhasználás stb.). A közelmúlt talán legígéretesebb fejleménye az adminisztratív adatbázisok és a kérdőíves felmérések összekapcsolása: a PSID és az HRS mintájában szereplő egyénekhez hozzákapcsolható teljes társadalombiztosítási és adóbefizetési történetük; a SHARE-adatokhoz sok országban hozzákapcsolhatók az egészségügyi ellátórendszerrel kapcsolatos, illetve a társadalombiztosítási adatok. Az innováció az adatok összekapcsolásában nem ismer határokat – hogy csak néhány példát említsünk: paneladatbázisok résztvevőiről gyakran gyűjtenek kiegészítő információkat, felkérlik őket kísérletekben való részvételekre, és kiegészítő felmérésekben megkeresik a kapcsolathálóikban szereplőket is (lásd például *Dohmen és szerzőtársai* [2010]). A kapcsolások révén az egyes adatbázisokban fellelhető információ értékét megsokszorozza egy másik adatbázis információ-tartalma, különösen akkor, ha legalább az egyikük panel jellegű.

Leírásunkban csak a felszínét érinthettük. Mivel napjaink empirikus eredményei nagyrészt panel jellegű adatbázisokra támaszkodnak, tömegével találhatunk a standard kétdimenziós mellett három-, illetve háromnál több dimenziós adatbázisokat, és ennél bonyolultabb relációs adatbázisokat is. Alapvető tendencia, hogy amikor lehet, elemzéseinkhez paneladatokat használunk. Ha más nem, legalább nagyobb mintaelemszámú adatbázishoz jutunk, mint ha pusztán keresztmetszeti vagy idősoros adatokkal dolgoznánk. A nagyobb adatbázisok a nagy mintaelemszám mellett nagyobb variációjú magyarázó változókhoz vezetnek, amelyek nemcsak pontosabb becsléseket tesznek lehetővé, hanem a magyarázó változók tágabb értelmezési tartományán való mérést is. A panel legfontosabb előnye mégis az, hogy szerencsés esetben megszüntethetjük – vagy legalábbis csökkenthetjük – az endogenitásból fakadó torzításokat, vagy éppen gazdagabb dinamikát vizsgálhatunk. Az ehhez kapcsolódó módszertani fejleményeket tárgyalja a következő rész.

### *Módszertani fejlemények*

Az elméleti panelökonometria részletes kidolgozottsága következtében egyre hangsúlyosabbá válik a már meglévő módszerek általánosítása és finomítása. Mint már említettük, egy-egy feltétel fellazítására gyakran egy adatbázis sarkallja az elméleti kutatókat, mely munka során igyekeznek a módosított becslőfüggvényeket identifikálhatóvá és konzisztenssé (azaz elemzési célokra használhatóvá) tenni. Egy-egy speciális adatstruktúra esetén az irodalomból már ismert modellek sok esetben ezekre a feladatokra nem képesek. Vegyük sorba tehát a panelökonometria főbb ágait és azok lényeges eredményeit!

Amikor panelökonometriáról beszélünk, az alkalmazók általában egyből a fix hatásokra gondolnak. Népszerűsége nem véletlen, hiszen a fix hatások segítenek kiszűrni az időállandó és/vagy individuumállandó nem megfigyelt tényezőket, melyek egyébként endogenitást okoznának. Vegyük a következő egyszerű példát! Koren Miklós és Csillag Márton az importált gépek hatását vizsgálja a dolgozók keresetére Magyarországon (*Koren–Csillag* [2011]). Ha cégek egy keresztmetszetét

vizsgáljuk, az importáló cégekben dolgozók kereseteit tudjuk összehasonlítani a nem importáló cégekben dolgozók kereseteivel. Az importáló cégeknél dolgozók átlagosan többet keresnek, de ebből természetesen nem következik, hogy az import okozza a magasabb kereseteket – a nem megfigyelhető heterogenitás valószínűleg korrelál az importálási eséllyel és a dolgozók összetételével is, ami endogenitást okoz. Természetesen megpróbálhatunk kontrollálni minél több ismérvre (a cégeknél dolgozó munkaerő-összetételére iskolai végzettség és kor szerint, a cégek iparági besorolására stb.), de sohasem lehetünk biztosak abban, hogy minden olyan tényezőre kontrolláltunk, amely az endogenitást okozza. Ha azonban paneladatunk van, a fix vállalathatás panelregressziójával azt vizsgálhatjuk meg, hogy ugyanannál a cégnél dolgozók keresete magasabb-e akkor, ha a cég importált gépet használ. Ezzel tulajdonképpen kontrollálni tudunk minden olyan – megfigyelhető és nem megfigyelhető – heterogenitásra a cégek között, amely időben állandó, kiküszöbölve ezek hatását a gépimport-keresetek összefüggésben.

Az elméleti ökonometria nyelvén ezt úgy fogalmazzuk meg, hogy a célt egy olyan becslés elérése jelenti, amely egy adott feltételrendszer esetén modellünket torzítatlanul vagy legalábbis konzisztensen becsüli. Konzisztencián azt értjük, hogy a becslés várható értékben közelít a valós paraméterértékhez, miközben növeljük az adatbázis méretét. Ez elsősorban elég elvontnak tűnik, de valójában egyszerűen azt jelenti, hogy ha elég sok megfigyelésünk van (köszönhetően a nagy adatbázisoknak), becslésünk várhatóan nagyon közel helyezkedik el ahhoz, amit meg szeretnénk határozni.

A kétdimenziós eset (például háztartások követése időben; országok követése időben) elméletét több évtizede kidolgozták, a közelmúlt fejleményei inkább a többdimenziós általánosításokkal foglalkoznak, de találunk olyan eredményeket is, amelyek a becslések robusztusságát javítják. A legkisebb négyzetek típusú becslés például köztudottan érzékeny a kiugró megfigyelésekre (*outlierekre*). Az ilyen kiugró értékek már egy keresztmetszeti adatbázisban is jelentős eltéréseket okozhatnak a valós paraméterértékektől, és ez a negatív hatás még erősebb egy paneladatbázis esetén. *Bramati–Croux* [2007] például bevezet egy ügyesen módosított *Within* becslést, amellyel a becsült paraméterek robusztussá tehetőek, azaz a becslés nem változik számottevően, ha az adatbázis apróbb részeit módosítjuk.<sup>2</sup> Ez különösen fontos lehet háztartásokból álló adatbázisok esetén, ahol az egyedek roppant heterogének. A tradicionális fixhatás-megközelítés alkalmazható nemlineáris és nem parametrikus modellek kezelésére is. A lineáris fixhatás-panelmodelleknél bevett lineáris transzformációk egy nemlineáris modell esetén a *Chamberlain* [1984] által taglalt feltételes maximum likelihood becsléssel helyettesíthetők [ezeknek a modelleknek egy részét korrelált véletlen hatás paneladatmodellekként (*Correlated Random Effects Panel Data Models*) is ismerik]. Ez a becslés általánosítható további nemlinearitások megengedésével, ami megsokszorozza a használhatóságát (lásd *Hoderlein és szerzőtársai* [2011]). Ezzel a fixhatás-modellek újabb osztálya nyílt meg, ahol az egyedek közötti, időben rögzített heterogenitás megengedhető, illetve kezelhető.

<sup>2</sup> Ismét egy kitűnő példa a gyakorlati megfigyelés sarkallta elméleti cikke.

Fix hatások nemcsak additívan, hanem interaktívan is szerepelhetnek egy modellben. Ilyen szerkezettel találkozhatunk például kereseti modellekben, ahol a munkavállalók kitaratása, hozzáértése és veleszületett képessége kombinálva hat a keresetre. Nem kell azonban messzire mennünk a makroökonómiában sem, ahol egy aggregált sokk és annak egyénekre vonatkozó hatása együttesen határoz meg egy ilyen interaktív hatást. Ezeknek a modelleknek az identifikálása és paramétereik konzisztens becslése speciális figyelmet igényel, ahogyan ezt Bai [2009] munkájában is láthatjuk. Érdekes (bár kissé bonyolult) Stock–Watson [2008] törekvése, hogy megszüntettesse a becsült variancia-kovariancia mátrix (gyakorlatilag a szórásbecslések) inkonzisztenciáját azokra az esetekre, amikor a vizsgált időintervallum rövid, és autokorreláció is jelen van.

A fixhatás-modellek keretében egyre nagyobb teret kapnak a dinamikus modellek. Amennyiben az adatainkból a függő változó autokorreláltságára következtünk, indokoltta válik annak késleltetett értékeit mint magyarázó változót a regressziós egyenletünkhöz adni. A fogyasztásom ebben a hónapban minden bizonnyal függ a múlt havi fogyasztásomtól. Egy cég munkakereslete nagyban függ a tavalyi munkakeresletétől. Ha a fenti kapcsolatokat figyelmen kívül hagyjuk, ugyanolyan hibát követünk el, mintha bármely másik fontos magyarázó változót hanyagolnánk el. Ha egy dinamikus modellt *Within* becsléssel akarunk becsülni (tehát jelen van valamilyen fixhatás-struktúra is), az így kapott becslőfüggvény rögzített  $T$ -re (az egyedi idősorok hossza)  $N$ -ben (a megfigyelt egyedek száma) inkonzisztens (ez Nickell [1981] munkájából a panelökonometria egy jól ismert eredménye). Ezt kiküszöbölendő számos megoldás született, melyek többnyire valamilyen ortogonalitási feltételrendszerrel használva állítanak fel  $N$ -ben konzisztenciát (a legkedveltebbek Anderson–Hsiao [1981] vagy Arellano–Bond [1991]). Everaert [2013] egészen más perspektívából vizsgálja ezt a kérdést. Egy olyan ortogonális transzformáción nyugvó becslést ad, amely ugyan rögzített  $T$ -re nem konzisztens, de a torzulás mértéke elhanyagolható. Ezzel még nem lennénk előrébb, de Monte-Carlo-szimulációkkal azt is megmutatja, hogy becslése jobb kis mintás tulajdonságokkal rendelkezik, mint más dinamikus becslések. Ez igen fontos, mivel e ponton előjön a panelökonometriával foglalkozók számára jól ismert átváltás (*trade-off*) a torzítottság és a precizitás (alacsony standard hiba) között. Hajlandók vagyunk-e szemet hunyni egy bizonyos fokú torzítottság felett, ha cserébe kisebb szórású, tehát precízebb becsléseket kaphatunk? A válasz természetesen függ az adott környezettől, a vizsgált kérdéstől, a torzítás és a variancia méretétől, de a gyakorlat azt mutatja, hogy gyakran fogadunk el, ha minimálisan is, de torzított becsléseket a nagyobb pontosság érdekében. A nagy adatbázisok ráirányítják a felhasználók figyelmét a régi igazságra, hogy nem a torzítatlanság vagy a konzisztencia az igazán fontos, hanem hogy a becsült paraméterek nagy valószínűséggel a keresett (de ismeretlen) paraméterek kellően kicsi környezetében legyenek. A módszertani kutatások ezen környezet szűkítésére és a valószínűség növelésére irányulnak, minél általánosabb feltételek mellett.

Hagyjuk el egy kis időre a lineáris modellek világát, és tekintsünk be a nemlineáris panelökonometria ugrásszerűen növekvő irodalmába! Kényelmes lineáris

modellekkel dolgozni, hiszen mind becslésük, mind értelmezésük viszonylag egyszerű. Sok jelenség leírásához azonban nem célszerű lineáris modelleket alkalmazni. Ha meg akarom határozni, hogy egy elítélt szabadulása után várhatóan mennyi idő múlva kerül vissza a börtönbe, vagy egy elbocsátott munkavállaló mennyi idő múlva jut újra munkához, kifinomultabb, nemlineáris függvényekre van szükség. Az említett példák az időtartammodellek (*duration analysis*) témakörébe tartoznak, amelyek leírásához a természetes modellkeret nemlineáris. Az olyan kérdések elemzéséhez, mint hogy mi az esélye, hogy egy kerületben szelektív hulladékgyűjtő épül, vagy hogy mennyivel nagyobb ez a valószínűség, ha a kerület gazdagabb, nagyobb, kisebb a bűnözés stb., valószínűségi modelleket írunk fel. A valószínűségi modell lineáris változata nem más, mint a lineáris regresszió, ami kényelmes, és bizonyos célra megfelelő lehet, más célokra azonban nem alkalmas, mert nincs garancia arra, hogy egy lineáris modell nulla és egy közötti számokat adna válaszul. A nemlineáris modellek becslésekor használt ortogonalitási feltételek a legtöbb esetben nem fejezhető ki zárt alakban, gyakran kell numerikus módszerekre (és ezzel intenzív számítógép-használatra) támaszkodni. Nem meglepő tehát, hogy a technológiai forradalommal együtt vált ez a terület is népszerűbbé.

A nemlineáris panelmodellek közül a bináris valószínűségi (probit és logit) modelleknek van a legnagyobb irodalma. Ezeknek különös a jelentőségük nagy adatbázisok esetén, mivel segítenek megérteni az adatok szerkezetét, vagyis hogy egyes mérések miért szerepelnek benne, mások pedig miért nem. Bármit is vizsgálunk, alapvető az első lépésben tisztázni a hiányzó adatok mibenlétét, az „adatlyukak” okait, ami leggyakrabban probit és logit modellek segítségével történik. A standard logit és probit modellek panelváltozatai nehezen becsülhetőek, és gyakran a statisztikai tulajdonságaik sem ismertek teljes mértékben. Akármilyen régi is ez az irodalom, a bináris valószínűségi modellek napjainkban is intenzív ökonometriaelméleti kutatás tárgyai, így meglehetősen nehéz kiválasztani a legfontosabb eredményeket. A következőkben említett cikkek, reméljük, átfogó képet adnak arról, hogy merre is halad és miért is sajátos ága ez a panelökonometriának. *Thomas* [2006] heterogén lineáris trendeket enged meg a logit fixhatás-modellben, majd szimulációs módszerekkel vizsgálja a feltételes logit becslésének kis mintaelemszámú tulajdonságait. Ezzel kezelhetővé tesz egy új modellcsoportot, amelyet például az orosz földbérleti szokások vizsgálatára is használtak. *Gu és szerzőtársai* [2009] érvelése alapján torzításhoz vezethet, ha elhanyagoljuk hibatagjaink heteroszkedaszticitását, így a javasolt bináris probit véletlenhatás-modellt komoly számításgépi bayesi technikákkal becsli. A már említett munka *Hoderlein és szerzőtársai* [2011] tollából kibővíti a bináris valószínűségi fixhatás-modelleket nem parametrikus komponensekkel, majd egy általánosított, szintén nem parametrikus módszereken alapuló becslést általánosít. De olyan nagy szerzők munkáival is találkozhatunk, mint *Chamberlain* [2010], aki bináris kimenetelek valószínűségének az identifikálhatóságával foglalkozik, vagy *Greene* [2004], aki rávilágít arra a tényre, hogy a sokak által kedvelt maximum likelihood becslés tulajdonságai fix hatás jelenlétében kevésbé ismertek a folytonos, nemlineáris modellek

[mint például a tobit vagy éppen a csonkolt (*truncated*) regressziós modell] esetén. Mivel ezek a tulajdonságok lényegesen eltérhetnek a standard bináris kiválasztási modellekétől, eredménye releváns mindazok számára, akik nem nélkülözhetik a fix hatásokat modelljeikből. Egyre több becsléskor hagyatkoznak a már szinte nélkülözhetetlen bayesi módszerekre, mely térnyerését szintén a számítógépek rohamos fejlődésének köszönheti.<sup>3</sup>

Az eddig taglalt nem standard (és újszerű) módszerek magukban hordozzák a jól ismert tesztek (legyen az Wald-, LM- vagy éppen LR-próba) módosítására való igényt.<sup>4</sup> A kidolgozott új tesztek működőképeseek egy lazább feltételrendszer esetén, és gyakran akár kedvezőbb kis mintás tulajdonságokkal is rendelkezhetnek. *Baltagi–Yang* [2013] alapján egyes LM-próbák véges mintaelemszámban érzékenyek lehetnek a félrespecifikálásra, és az általuk megjelölt LM-próba jól teljesít mind aszimptotikusan, mind kis mintaelemszámban. Mivel a félrespecifikálás gyakran okoz fejfájást a kutatóknak, igazi megkönnyebbülést jelent, hogy legalább a tesztelés során kevésbé kell tartanunk tőle. Találkozhatunk a híres Breusch–Pagan-féle LM-próba (*Breusch–Pagan* [1980]) torzításigazított (*bias-adjusted*) verziójával *Pesaran és szerzőtársai* [2008] munkájában, illetve különböző tesztekkel stacionaritás vizsgálatára (például *Hadri és szerzőtársai* [2012] tesztje rögzített  $T$  esetén). Napjaink modelljei annyira egyedülállóak tudnak lenni, hogy nemritkán a modellekkel együtt tesztet is közölnek az írók. Ez nagyban segíti az eredmények gyakorlati felhasználhatóságát, mivel, mint már említettük, a standard próbák gyakran nem érvényesek egy speciális környezetben. E nélkül gyakorlatilag önálló kutatásként kellene a felhasználóknak elkészíteni a szükséges tesztek.

Az említett nagyobb témákon kívül természetesen számos más irányba is halad a panelökonometria elméleti kutatása. Érdeemes kiemelni a (nagyon) nagy paneleket, illetve a kezelésükre irányuló törekvéseket. E panelek előnye, hogy pontosabban érvényesülnek rájuk az aszimptotikus tulajdonságok, így sok apró finomító lépést tehetünk egy jobb tulajdonságú vagy egyszerűen csak hatásosabb módszer kifejlesztéséhez. Ilyen törekvés például a *bootstrap* szimuláció finomítása (lásd *Kapetanios* [2008]). Az adatbázisok egyre nagyobbak és egyre gazdagabbak, ám a részleges adathiány legalább annyira probléma esetükben, mint a hagyományosabb paneladatbázisokban. Minél nagyobb az adatbázis, annál valószínűbb, hogy hiányos, így annak, aki napjaink nagy adatbázisait elemzi, különösen fontos számolnia ezzel a problémával. Mivel egy nem teljes adatbázis sok esetben torzított becslésekhez vezet, és nem mindig engedhetjük meg a luxust, hogy elhagyjuk az adatbázis „nem teljes” részét, egyre több munka irányul a hiányos panelek felé is. Egy kidolgozott módszer mit sem ér, ha az a szinte már csak könyvekből ismert teljes adatbázisokon alkalmazható. Köztudott például, hogy a fix hatásos

<sup>3</sup> Dióhéjban, kiindulunk egy közös *prior* eloszlásból, amelyről azt gondoljuk, hogy a paramétereink valós eloszlásához közel állhat, majd rendelkezésre álló adatainkból a bayesi szabály (általában többszöri) alkalmazásával, numerikus módszerekkel megkapjuk az a *posterior* eloszlást, amely bizonyos regularitási feltételek mellett tetszőlegesen közel állhat a paraméterek valós eloszlásához.

<sup>4</sup> A legtöbb teszt a fenti három kategória valamelyikébe sorolható: Wald-teszt, Lagrange-multiplikátor (LM) vagy likelihood arány (LR).

panelmodellek kezelésére használt vetítések nem feltétlen használhatók hiányos adatbázis esetén, ahogy ezt *Abrevaya* [2013] is boncolgatja (és javasolja egy alternatív, általánosított vetítés használatát).

E néhány bekezdés jól illusztrálja, hogy jelenleg merre is halad a panelökonometria. Az elmúlt egy-két évtized eredményeit nézve, látható, hogy gyorsan el lehet veszni a részletekben, így a fenti felsorolás inkább irányadóként szolgál, mintsem a terület eredményeinek egy kimerítő gyűjteményét adná. A főbb folyamatok és irányok azonban, reméljük, jól elkülöníthetők.

### *Néhány kiragadott alkalmazás*

Végül vizsgáljunk meg kicsit részletesebben néhány olyan alkalmazást, amelyek jelentősek akár a panelökonometria fejlődése, akár a vizsgált tartalmi kérdés fontossága szempontjából! Az áttekintés rövid és szubjektív, behatárolják a terjedelmi korlátok és e cikk szerzőinek az érdeklődési köre. Először a nemzetközi kereskedelem példáját vizsgáljuk, ami azért tanulságos, mert sok olyan kérdéssel foglalkozik, amelyek természetes módon vezetnek igen bonyolult panelmodellekhez. Ezután megemlítünk néhány makroökonómiai, piacszerkezeti és munka-gazdaságtani alkalmazást is.

A nemzetközi kereskedelem empirikus elemzése ma már szinte kizárólag paneladatbázisokra épül. Két ország kapcsolatát egyetlen időpontban vizsgáló keresztmetszeti elemzések valószínűleg torz és inkonzisztens választ adnának a kérdéseinkre, és nem engednék meg a hatások dinamikájának elemzését sem. Nem kevés példát láthatunk régi keresztmetszeti eredmények paneladatbázison való újrabecslésére, ami gyökeresen más konklúzióhoz vezet (*de Mello* [1999], *Rodgers* [2001], *Rose* [2004]). Sokan használják a panel-adatstruktúra nyújtotta lehetőségeket bilaterális kereskedelem modellezésére és tesztelésére: *Disdier–Head* [2008] szisztematikus bizonyítást ad a távolság negatív szerepére, *Fieler* [2011] bővíti a híres gravitációs modellt, amely így már számításba veszi az egy főre jutó jövedelem eloszlását is. Számptalan példát találunk olyan modellekre, amelyek a termelékenység, technológiai fejlődést, export-importot, működőtőkét, országok közötti egyenlőtlenséget magyarázzák leggyakrabban termékszintig lemenő paneladatbázisok használatával. Ilyenek például *Helpman és szerzőtársai* [2008], [2010], *Lentz–Mortensen* [2008], vagy *Eaton és szerzőtársai* [2011].

A makroökonómiában az idősorok vizsgálata a jellemző, de a fontos panelalkalmazások közé tartozik például *Dynan* [2000] a lakossági fogyasztási szokásformálásról vagy *Bils–Klenow* [2004] az árak dinamikájáról. A piacszerkezet területén széles körben fontos eredményként tartják számon a *Bloom és szerzőtársai* [2012] tanulmányt, amely az Európában működő amerikai vállalatokon keresztül világít rá az információs technológia fontosságára, a *Grennan* [2013] cikket, amely paneladatbázis empirikus vizsgálatával igazolja a középkori árdiszkriminációt, vagy éppen az *Einav és szerzőtársai* [2013] tanulmányt, amely dolgozók egészségügyi biztosításával kapcsolatban mutat ki erkölcsi kockázatot. A munkagazdaságtanban

a rengeteg új panelalkalmazás közül példaként említhetjük az emberitőke-elmélet többtényezős és dinamikus modelljének elemzését (*Cunha–Heckman* [2008]) vagy a jobb munkapiaci körülmények indukálta migrációs döntések dinamikus elemzését (*Kennan–Walker* [2011]).

Talán nem túlzás kijelenteni, hogy a paneladatok használata nagy lehetőségeket hordoz magában az empirikus közgazdasági elemzés mindent területén. Gyakran korai eredményeket gondolnak újra, és próbálnak tovább erősíteni paneladatok segítségével. A paneladatok nemcsak ismert problémák kezelésére adhatnak lehetőséget, hanem teljesen új kérdések megválaszolását teszik lehetővé. Akármilyen céllal alkalmazzuk is, a panelökonometria gyakran megbízhatóbb, érvényesebb és kifinomultabb elemzést tesz lehetővé.

## Összefoglalás és kitekintés

Láthattuk, hogy a panel jellegű adatok elérhetőségének javulása komolyan befolyásolta a panelökonometria fejlődését. Remek példával szolgál egy alulról építkező tudományág fejlődésére. Klasszikusan műhelytanulmányok, PhD-dolgozatok, gyakorta új adatbázisokra támaszkodó cikkek vonalán indul, amelyeket összefoglaló könyvek és külön, e témára célzott konferenciák, konferenciasorozatok követnek. Ezek később visszahatnak újabb adatbázisok keletkezésére, gyűjtésére, összekapcsolására és így tovább. Láthatjuk tehát, hogy a panelökonometria hatalmas irodalma hamar magával ragadja az olvasót. Gyakorlati használhatósága, egyre növekvő és fejlődő teoretikus háttere joggal teszi napjaink egyik legimpozánsabb területévé.

A Nobel-díjas Granger szavaival élve: az utóbbi évek az adatforradalom korszakváltásáról szóltak: az adatszűkéből egyre inkább az adatok túltengésének állapotába jutunk (*Granger* [1998]). Ezzel párhuzamosan ugrásszerűen fejlődik az adattárolás és az adatelemzés technikai háttere is. A panelökonometria első lépései óta az adatok tárolási kapacitása és a számolási sebesség milliós nagyságrenddel nőtt.

Nem vállalunk nagy kockázatot azzal, ha a közeljövőben ezeknek a folyamatoknak a folytatódását (és fokozódását) vetítjük előre. A különböző szenzorok fejlődésével elképesztő ütemben bővülnek a rendelkezésünkre álló mérések, aminek hatását a gazdasági elemzésekre még nem tudjuk felmérni. Jobban látható annak a hatása, hogy a számítógépeknek a gazdasági élet minden területén való megjelenésével és elterjedésével automatikusan gyűjtünk korábban elképzelhetetlen mennyiségű adatot közgazdasági elemzésekhez hagyományosan használt területeken (gondoljunk az üzletekben használt elektronikus árucikk-leolvasó rendszerekre vagy egyes városok metróhálózatába való chip alapú beléptetésre, hogy csak két teljesen hétköznapi példát említsünk). Az adatmennyiség robbanásszerű bővülésének harmadik motorja az internet. Az adatok szűkösségében felnőtt nemzedékek számára elképzelhetetlen mennyiségű digitális információ van a napi több mint hárommilliárd internetes keresést kiszolgáló Google tulajdonában (a Google többi termékét, mint például a több száz milliónyi email-fiókban tárolt leveleket most nem is számolva). Ugyannyire hihetetlen mértékű adat ülhet a – Wikipedia szerint évente 74 milliárd



dollárnyi – internetes vásárlást lebonyolító Amazon szerverein, vagy a több mint egymilliárd felhasználóval rendelkező Facebook tárolóiban.

Az adatforradalom (jelenleg divatos kifejezéssel élve a *big data* forradalom) nyomán egyre több, hatalmas méretű, sokszor igen bonyolult szerkezetű („rendszeretlen”) és hibákkal, adathiányokkal terhelt („zajos”) adat áll az elemzők rendelkezésére. A számítástechnika irányából érkező elemzők különböző nem ökonometriai, sokszor mesterséges intelligenciára (gépi tanulásra) épülő algoritmusok alapján próbálnak információt nyerni ezekből az adatokból. Meg vagyunk azonban győződve arról, hogy a panelökonometria eszköztára és megközelítési módja sokat adhat az ilyen adatok elemzéséhez. Természetesen nem vagyunk egyedül ezzel a vélekedéssel: nemcsak az ökonometria művelői gondolkodnak hasonlóan, de néhányan az informatikai iparágban is sürgetik a gépi tanulási módszerek ötvözését a hagyományos ökonometriai eszköztárral (lásd *Varian* [2014]). Az pedig biztosra vehető, hogy az adatforradalom nagy hatással lesz a panelökonometria fejlődésére. A panelökonometria ismerete mindazok számára megkerülhetetlenné válik, akik követni szeretnék a jövő empirikus közgazdaság-tudományi fejlődését – akár elemzőként, akár az elemzések felhasználóiként, olvasóiként.

### Hivatkozások

- ABREVAYA, J. [2013]: The projection approach for unbalanced panel data. *The Econometrics Journal*, Vol. 16. No. 2. 161–178 o.
- ADAMS, P.–HURD, M. D.–MCFADDEN, D.–MERRILL, A.–RIBERIO, T. [2003]: Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status. *Journal of Econometrics*, Vol. 112. No. 1. 3–56 o.
- AIRY, SIR G. B. [1861]: *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations*. Macmillan, Cambridge–London.
- ANDERSON, T. W.–HSIAO, C [1981]: Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*, Vol. 76. 598–606. o.
- ARELLANO, M.–BOND, S. [1991]: Some Tests of Specification for Panel Carlo Application to Data: Evidence and an Employment Equations. *Review of Economic Studies*, Wiley Blackwell, Vol. 58. No. 2. 277–297. o.
- BAGHESTANI, H. [2011]: Federal Reserve and Private Forecasts of Growth in Investment. *Journal of Economics and Business*, Vol. 63. No. 4. 290–305. o.
- BAI, J. [2009]: Panel Data Models with Interactive Fixed Effects. *Econometrica*, Vol. 77. No. 4. 1229–1279. o.
- BALESTRA, P.–NERLOVE, M. [1966]: Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica*, Vol. 34. No. 4. 585–612. o.
- BALTAGI, B. H. [1995]: *Econometric Analysis of Panel Data*. Wiley, New York.
- BALTAGI, B. H.–YANG, Z. [2013]: Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal*, Vol. 16. No. 1. 103–134. o.
- BARSKY, R. B.–KIMBALL, M. S.–JUSTER, F. T.–SHAPIRO, M. D. [1997]: Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement survey. *The Quarterly Journal of Economics*, Vol. 112. No. 2. 537–579. o.

- BAUER, A.–EISENBEIS, R. A.–WAGGONER, D. F.–ZHA, T. [2003]: Forecast Evaluation with Cross-Sectional Data: The Blue Chip Surveys. *Economic Review*, Vol. 88. 17–31. o.
- BILS, M.–KLENOW, P. J. [2004]: Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, Vol. 112. No. 5. 947–985.
- BJÖRKLUND, A.–JÄNTTI, M.–SOLON, G. [2007]: Nature and Nurture in the Intergenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents. *The B.E. Journal of Economic Analysis and Policy*, Vol. 7. No. 2. 1–23. o.
- BLACK, S. E.–DEVEREUX, P. J.–SALVANES, K. G. [2005]: The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *The Quarterly Journal of Economics*, Vol. 120. No. 2. 669–700. o.
- BLANCHARD, O. J.–DIAMOND, P. [1990]: The Cyclical Behavior of the Gross Flows of U.S. Workers. *Brookings Papers on Economic Activity*, 2. 87–155. o.
- BLOOM, N.–SADUN, R.–REENEN, J. VAN [2012]: Americans Do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review*, Vol. 102. No. 1. 167–201. o.
- BRAMATI, M. C.–CROUX, C. [2007]: Robust Estimators for the Fixed Effects Panel Data Model. *The Econometrics Journal*, Vol. 10. No. 3. 521–540. o.
- BREUSCH, T. S.–PAGAN, A. R. [1980]: The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics. *Review of Economic Studies*, Vol. 47. No. 1. 239–253. o.
- CALVET, L. E.–CAMPBELL J. Y.–SODINI, P. [2008]: Fight or Flight? Portfolio Rebalancing by Individual Investors. *The Quarterly Journal of Economics*, Vol. 124. No. 1. 301–348. o.
- CARTER, K. N.–CRONIN, M.–BLAKELY, T.–HAYWARD, M.–RICHARDSON, K. [2010]: Cohort Profile: Survey of Families, Income and Employment (SoFIE) and Health Extension (SoFIE-health). *International Journal of Epidemiology*, Vol. 39. No. 3. 653–659. o.
- CHAMBERLAIN, G. [1984]: Panel data. Megjelent: *Griliches, Z.–Intriligator, M. D.* (szerk.): *Handbook of Econometrics*, Vol. 2. Elsevier Science Publishers, 1247–1318. o.
- CHAMBERLAIN, G. [2010]: Binary Response Models for Panel Data: Identification and Information. *Econometrica*, Vol. 78. No. 1. 159–168. o.
- CHAMBERLAIN, G.–GRILICHES, Z. [1975]: Unobservables with a Variance-Components Structure. Ability, Schooling, and the Economic Success of Brothers. *International Economic Review*, Vol. 16. No. 2. 422–449. o.
- CHOY, K. M.–LEONG, K.–TAY, A. S. [2006]: Non-fundamental Expectations and Economic Fluctuations: Evidence from Professional Forecasts. *Journal of Macroeconomics*, Vol. 28. No. 2. 446–460. o.
- COAD, A.–BINDER, M. [2014]: Causal Linkages between Work and Life Satisfaction and Their Determinants in a Structural VAR Approach. *Economics Letters*, Vol. 124. No. 2. 263–268. o.
- COILE, C. C.–LEVINE, P. B. [2006]: Bulls, Bears, and Retirement Behavior. *Industrial and Labor Relations Review*, Vol. 59. No. 3. 408–429. o.
- CROUSHORE, D. [1993]: Introducing: The Survey of Professional Forecasters. *Federal Reserve Bank of Philadelphia Business Review*, 3–13. o.
- CUNHA, F.–HECKMAN, J. J. [2008]: Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, Vol. 43. No. 4. 738–782. o.
- DE MELLO, L. R. [1999]: Foreign Direct Investment-Led Growth: Evidence from Time Series and Panel Data. *Oxford Economic Papers*, Vol. 51. No. 1. 133–151. o.

- DISDIER, A.-C.–HEAD, K. [2008]: The Puzzling Persistence of the Distance Effect on Bilateral trade. *The Review of Economics and Statistics*, Vol. 90. No. 1. 37–48. o.
- DIVÉNYI JÁNOS–KÉZDI GÁBOR [2013]: Low Employment among the 50+ Population in Hungary: the Role of Incentives, Health and Cognitive Capacities. Megjelent: *Börsch-Supan, A.–Brandt M.–Litwin, H. Weber, G. (szerk.): Active Ageing and Solidarity between Generations in Europe: First Results from SHARE After the Economic Crisis*. De Gruyter, Berlin (77–90. o.).
- DOHMEN, T.–FALK, A.–HUFFMAN, D.–SUNDE, U. [2010]: Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review*, Vol. 100. No. 3. 1238–1260. o.
- DYNAN, K. E. [2000]: Habit Formation in Consumer Preferences: Evidence from Panel Data. *American Economic Review*, Vol. 90. No. 3. 391–406. o.
- EATON, J.–KORTUM, S.–KRAMARZ, F. [2011]: An Anatomy of International Trade: Evidence from French Firms. *Econometrica*, Vol. 79. No. 5. 1453–1498. o.
- EINAV, L.–FINKELSTEIN, A.–RYAN, S. P.–SCHRIMPF, P.–CULLEN, M. R. [2013]: Selection on moral hazard in health insurance. *American Economic Review*, Vol. 103. No. 1. 178–219. o.
- EISENHART, C. [1947]: The Assumptions Underlying the Analysis of Variance. *Biometrics*, Vol. 3. No. 1. 1–21. o.
- EVANGELOS, F.–PETERS, E. [1998]: Survey Attrition and Schooling Choices. *The Journal of Human Resources*, Vol. 33. No. 2. 531–554. o.
- EVERAERT, G. [2013]: Orthogonal to Backward Mean Transformation for Dynamic Panel Data Models. *The Econometrics Journal*, Vol. 16. No. 2. 179–221. o.
- FIELDER, A. C. [2011]: Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation. *Econometrica*, Vol. 79. 1069–1101. o.
- FISHER, R. A. [1918a]: The Causes of Human Variability. *The Eugenics Review*, Vol. 10. No. 4. 213–220. o.
- FISHER, R. A. [1918b]: The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* Cambridge University Press, Vol. 52. No. 2. 399–433. o.
- FISHER, R. A. [1925]: *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh–London.
- GILES, P. [1999]: Overview of the Survey of Labour and Income Dynamics (SLID). *Survey of Labour and Income Dynamics*, Statistics Canada.
- GOTTSCHALK, P.–RUGGLES, P. [1994]: Using the Panel Study of Income Dynamics to study poverty and welfare dynamics. PSID Board Commissioned Papers: Design Options for Next Funding Cycle. Boston College; Joint Economic Committee, United States Congress.
- GRANGER, C. W. J. [1998]: Extracting Information from Mega-Panels and High-Frequency Data. *Statistica Neerlandica*, Vol. 52. No. 3. 258–272. o.
- GREENE, W. [2004]: The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects. *The Econometrics Journal*, Vol. 7. 98–119. o.
- GRENNAN, M. [2013]: Price Discrimination and Bargaining: Empirical Evidence from Medical Devices. *American Economic Review*, Vol. 103. No. 1. 145–177. o.
- GROGGER, J. [1995]: The Effect of Arrests on the Employment and Earnings of Young Men. *The Quarterly Journal of Economics*, Vol. 110. No. 1. 51–71 o.
- GU, Y. G.–FIEBIG, D. G.–CRIPPS, E.–KOHNS, R. [2009]: Bayesian Estimation of a Random Effects Heteroscedastic Probit Model. *The Econometrics Journal*, Vol. 12. No. 2. 324–339. o.

- GUSH, K. – SCOTT, J. – LAURIE, H. [2013]: Households' Responses to Spousal Job Loss: 'All Change' or 'Carry on as Usual'? Understanding Society Working Paper Series, 2013-04.
- HADRI, K. – LARSSON, R. – RAO, Y. [2012]: Testing for Stationarity with a Break in Panels where the Time Dimension is Finite. *Bulletin of Economic Research*, Vol. 64. No. 1. S123–148. o.
- HAJDU, TAMÁS – KERTESI GÁBOR – KÉZDI GÁBOR [2014]: Roma fiatalok a középiskolában. Megjelent: *Kolosi Tamás–Tóth István György* (szerk.): Társadalmi Riport. 2014. Tárki, megjelenés alatt.
- HAUSMAN, J. A. [1978]: Specification Tests in Econometrics. *Econometrica*, Vol. 46. No. 6. 1251–1271. o.
- HECKMAN, J. J. – STIXRUD, J. – URZUA, S. [2006]: The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, Vol. 24. No. 3. 411–482 o.
- HELPMAN, E. – MELITZ, M. – RUBINSTEIN, Y. [2008]: Estimating Trade Flows: Trading Partners and Trading Volumes. *The Quarterly Journal of Economics*, MIT Press, Vol. 123. No. 2. 441–487. o.
- HELPMAN, E. – ITSCHKOKI, O. – REDDING, S. [2010]: Inequality and Unemployment in a Global Economy. *Econometrica*, Vol. 78. No. 4. 1239–1283. o.
- HILDRETH, C. [1950]: Combining Cross Section Data and Time Series. Cowles Commission Discussion Paper, Vol. St347. Letölthető a <http://dido.wss.yale.edu/P/ccdp/index.htm#stat> oldalról.
- HILL, M. S. [1992]: *The Panel Study of Income Dynamics: A User's Guide*. Sage Publications, Newbury Park, CA.
- HOCH, I. [1958]: Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function. *Econometrica*, Vol. 26. No. 4. 566–578. o.
- HOCH, I. [1962]: Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data. *Econometrica*, Vol. 30. No. 1. 34–53. o.
- HODERLEIN, S. – MAMMEN, E. – YU, K. [2011]: Non-parametric Models in Binary Choice Fixed Effects Panel Data. *The Econometrics Journal*, Vol. 14. No. 3. 351–367. o.
- HSIAO, C. [1986]: *Analysis of Panel Data*. Cambridge University Press, Cambridge.
- KAPETANIOS, G. [2008]: A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal*, Vol. 11. No. 2. 377–395. o.
- KEANE, M. P. – WOLPIN, K. I. [1997]: The Career Decisions of Young Men. *The Journal of Political Economy*. Vol. 105. No. 3. 473–522. o.
- KENNAN, J. – WALKER J. S. [2011]: The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, Vol. 79. No. 1. 211–251. o.
- KLIPS [2007]: Korean Labor and Income Panel Study (KLIPS) 1998–2006. 1–9. hullám, Korean Labor Institute, Korea Employment Information Service. (KEIS) Szöul, Korea, <http://eng.keis.or.kr/eng/project/survey/laborer.jsp>.
- KOREN MIKLÓS – CSILLAG MÁRTON [2011]: Machines and Machinists: Capital-Skill Complementarity from an International Trade Perspective. CeFiG Working Papers, No. 13.
- LEMOLA, S. – RICHTER, D [2014]: Institutional Rearing is Associated with Lower General Life Satisfaction in Adulthood. *Journal of Research in Personality*, Vol. 48. No. 1. 93–97. o.
- LENTZ, R. – MORTENSEN, D. T. [2008]: An Empirical Model of Growth through Product Innovation. *Econometrica*, Vol. 76. No. 6. 1317–1373. o.
- LILLARD, L. A. – WILLIS, R. J. [1978]: Dynamic Aspects of Earning Mobility. *Econometrica*, Vol. 46. No. 5. 985–1012. o.

- MADDALA, G. S. [1987]: Limited Dependent Variable Models Using Panel Data. *The Journal of Human Resources*, Vol. 22. 307–338. o.
- MARSCHAK, J.–ANDREWS, W. H. [1944]: Random Simultaneous Equations and the Theory of Production. *Econometrica*, Vol. 12. No. 3–4. 143–205. o.
- MÁTYÁS LÁSZLÓ–SEVESTRE, P. [1995]: *The Econometrics of Panel Data*. 2. javított kiadás, Kluwer Academic Publishers, Dordrecht.
- MC FALL, S.–BURTON, J.–JACKLE, A.–LYNN, P.–UHRIG, N. [2013]: *Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1-5, User Manual*. University of Essex, Colchester, 1–66. o.
- MCGARRY, K. [1999]: Inter Vivos Transfers and intended bequests. *Journal of Public Economics*. Vol. 73. No. 3. 321–351. o.
- MCGARRY, K. [2003]: Do Changes in Health Affect Retirement Expectations? *Journal of Human Resources*, Vol. 39. No. 3. 624–648. o.
- MUNDLAK, Y. [1961]: Empirical Production Functions Free of Management Bias. *Journal of Farm Economics*, Vol. 43. No. 1. 44–56. o.
- NANDI, A.–PLATT, L. [2013]: Britishness and Identity Assimilation among the UK's Minority and Majority ethnic groups. *Understanding Society, Working Paper Series*, Vol. 2013-08.
- NEAL, D.–JOHNSON, W. R. [1996]: The Role of Premarket Factors in Black-White Wage Differences. *The Journal of Political Economy*, Vol. 104. No. 5. 869–895. o.
- NERLOVE, M. [2005]: *Essays in Panel Data Econometrics*. Cambridge University Press, Cambridge.
- NICKELL, S. [1981]: Biases in Dynamic Models with Fixed Effects. *Econometrica*, Vol. 49. No. 6. 1417–1426. o.
- PESARAN, M. H.–ULLAH A.–YAMAGATA, T. [2008]: A Bias-Adjusted LM Test of Error Cross-Section Independence. *Econometrics Journal*, Vol. 11. No. 1. 105–127. o.
- RAO, C. R. [1952]: *Advanced Statistical Methods in Biometric Research*. Wiley, New York.
- RODGERS, J. R. [2001]: A Panel-Data Study of the Effect of Student Attendance on University Performance. *Australian Journal of Education*, Vol. 45. No. 3. 284–295. o.
- ROSE, A. K. [2004]: Do We Really Know That the WTO Increases Trade? *American Economic Review*, Vol. 94. No. 1. 98–114. o.
- SCHEFFÉ, H. [1956]: Alternative Models for the Analysis of Variance. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, Vol. 27. No. 2. 251–271. o.
- SCHEFFÉ, H. [1959]: *The Analysis of Variance*. John Wiley and Sons, New York.
- SIK ENDRE–TÓTH ISTVÁN GYÖRGY (szerk.) [1998]: *Magyar Háztartás Panel – Műhelytanulmányok 9. Jelentés a Magyar Háztartás Panel 6. hullámának eredményeiről*. Társi, Budapest.
- SMITH, J. P. [1994]: *New Directions in Socioeconomic Research on Aging*. Megjelent: *Abel, R. P.–Gift, H.–Ory, M.* (szerk.): *Aging and Quality of Life*. Springer, New York, NY. 15. fejezet, 275–294. o.
- SMITH, J. P. [1999]: The Size Distribution of Wealth in the United States: A Comparison among Recent Household Surveys. Megjelent: *Smith, J. P.–Willis, R. J.* (szerk.): *Wealth, Work, and Health: Innovations in Measurement in the Social Sciences*. University of Michigan Press, 209–232. o.
- SPÉDER ZSOLT [2001]: *Poverty Dynamics in Hungary during the Transformation: Essays in Economy and Society*. Megjelent: *Meusberger, P.–Jöns, H.* (szerk.): *Transformation in Hungary*. Physica-Verlag, Heidelberg, 225–248. o.
- STOCK, J. H.–WATSON, M. W. [2008]: Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression. *Econometrica*, Vol. 76. No. 1. 155–174. o.

- THOMAS, A. [2006]: Consistent Estimation of Binary-Choice Panel Data Models with Heterogeneous Linear Trends. *Econometrics Journal*, Vol. 9. No. 2. 177–195. o.
- TREWIN, D. [2005]: History of the monthly Labour Force Survey. 2005 Year Book Australia. Canberra: Australian Bureau of Statistics. 212–214. o.
- VARIAN, H. R. [2014]: Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, Vol. 28. No. 2. 3–28. o.
- WAGNER, G. G.–FRICK, J. R.–SCHUPP, J. [2007]: The German Socio-Economic Panel Study (SOEP)–Scope, Evolution and Enhancements. *Schmollers Jahrbuch* Vol. 127. No. 1. 139–169. o.
- WOLFF, E. N. [1998]: Recent Trends in the Size Distribution of Household Wealth. *Journal of Economic Perspectives*, American Economic Association, Vol. 12. No. 3. 131–150. o.
- WOODEN, M.–FREIDIN, S.–WATSON, N. [2002]: The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1. *Australian Economic Review*, Vol. 35. No. 3. 339–348. o.