
Műhely

Vincze János,

a Budapesti Corvinus Egyetem egyetemi tanára és a Magyar Tudományos Akadémia Közgazdaság- és Regionális Tudományi Kutatóközpont Közgazdaságtudományi Intézetének tudományos tanácsadója.

E-mail: janos.vincze@uni-corvinus.hu

Adat és információ – A statisztikai rendszerek feladatairól

A *Statisztikai Szemle* 2015. évi 7. számában megjelent cikkében *Bod Péter Ákos* elsősorban gazdaságpolitikai szempontból vizsgálja azt, hogy milyen újszerű követelmények merülnek fel a statisztikai rendszerekkel szemben (*Bod [2015]*). A közelmúlt világgazdasági eseményei által kiváltott igényeken kívül ír számos olyan – elsősorban a GDP¹-statisztikákkal kapcsolatos – problémáról, amelyek már régóta napirenden vannak. Cikke végén kijelenti: „a statisztika előállítóitól és elemzőitől többet és mást vár el a társadalom.” Írásomban arra keresem a választ, hogy szerintem mit kellene elvárnia a társadalomnak a hivatalos statisztikai rendszerektől, érzésem szerint ugyanis nem azt várja, amit kellene. Mindenekelőtt szeretném felhívni a figyelmet egy fontos különbségre, nevezetesen az adat és az információ közti eltérésre. Ezt követően a gazdaságstatisztika hagyományos szemléletével foglalkozom (aggregációs paradigma), majd annak hátulütőivel. A következőkben vizsgálni szándékozom az ún. big data-forradalom hatását a kérdéskör egészére. Végezetül az előzőkben kifejtetteket egy példán illusztrálom, majd összefoglalom a tanulságokat, és konkrét elképzeléseimet.

Szupermarketek, ahol a vásárló személyét a bankkártya (majdnem) tökéletesen azonosítja, rengeteg adatot tárolnak bevásárlásainkról. Szerintük ezt az adattömeget fel is használják, amennyiben képesek (ügyesen felhasználva a máshonnan szerzett adatokat) ténylegesen identifikálni az egyes vevőket. A sajtóban több történet is forog, amelyek szerint ott jobban ismernek minket, mint mi saját magunkat.² Úgy tűnik a szuper-

¹ GDP (gross domestic product): bruttó nemzeti termék.

² A legtöbbet idézett történet valószínűleg annak a terhes diáklánynak az esete, akinek a szülei úgy értesültek leányuk állapotáról, hogy pelenkavásárlásra jogosító kuponokat hozott a posta, még mielőtt arra szükség lett volna. Az adatbányász kereskedő „indirekt” jelekből következtetett. <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>.

marketek mint kereskedelmi vállalkozások számára ezek az adatok információk, azaz nem csak „ott vannak”, hanem használni is tudják ezeket céljaik elérésére.

1. Adat, információ és tudás

A hétköznapi nyelvhasználatban nem szokás megkülönböztetni az adatot az információtól, de néha célszerű, és a következőkben én is ezt fogom tenni. Szinte adottnak vesszük, hogy az adatokhoz tudás is társul, az adatok felhasználásának tudása, de ennek mértéke adatonként és felhasználónként is változik.³

Információ lenne-e a vásárlási adatbázis egy gazdaságpolitikusnak? Valószínűleg csekély mértékben, mivel sem a céljai, sem a képzettsége (pontosabban a háttérben működő, bázisul szolgáló elemzők képzettsége) nem tennék lehetővé, hogy használni tudja az adatokat. Mi lett volna a helyzet, ha létezett volna egy hasonló adatbázis száz évvel ezelőtt? Akkor valószínűleg a kereskedelmi vállalat sem igen tudott volna mit kezdeni a hatalmas adatmennyiséggel, ami papíron lett volna hozzáférhető, és tárolása is jelentős erőforrásokat kötött volna le. Lehet, hogy akkor is voltak szakemberek, akik elvben fel tudták volna használni a sok adatot, de sem a számítási, sem az adatátviteli kapacitások nem voltak azon a szinten, hogy gyakorlatilag bármit is kezdjenek egy nagy adatbázissal. A vállalat elemzői, ha egyáltalán alkalmaztak kvantitatív piacelemzési módszereket, akkor is csak kis mintán dolgozhattak, a meglévő adatok nagy részét el kellett dobniuk.

Nyilvánvalóan éles az ellentét a jelen és a múlt között, az információs technológiák fejlődése a big data-forradalomhoz vezetett, azaz ahhoz, hogy a nagy adatbázisokat ma már lehet és érdemes is fenntartani, mert valóban információt szolgáltatnak, gyakorlati és tudományos célokra is használhatók. Manapság igen nagy a gazdaságpolitikusok számára hozzáférhető adatok mennyisége is, de kérdés, hogy ezek információtartalma is ugyanolyan mértékben nőtt-e. Vagyis gazdaságpolitikai, hangsúlyozottan nem üzleti, döntéshozatal esetén a sok adat sok információt jelent-e, rendelkezünk-e azzal a tudással, ami az adatokat információvá teszi? Egy széles körben elterjedt nézet szerint például a kereskedelem számára ezt a tudást az ún. adatbányász vagy statisztikai tanulási módszerek széles körű alkalmazhatósága és rohamos fejlődése tette lehetővé (Ayers [2007]). Közgazdaságtani, illetve makroökonómiai problémákra azonban ezeket a módszereket még nemigen használták. Más út is el-

³ Az információ szó hétköznapi használatára példa a következő: „Megfulladunk az információtól, de szomjazunk a tudásra.” (Hastie–Tibshirani–Friedman [2001] 115–163. old.) Az én szóhasználatomban információ helyett adatokat kellett volna írni. A tudás az, ami az adatokat információvá alakítja. Ez a weboldal http://www.diffen.com/difference/Data_vs_Information sok érdekeset ír a két kifejezés közti jelentéskülönbségről.

képzeltető, hiszen a közgazdasági elmélet fejlődése is elvezethetett volna oda, hogy az adatok információvá váljanak. A közgazdasági elmélet fejlődésével kapcsolatban meglehetősen szélsőséges nézetek vannak forgalomban, és itt ezekkel nem akarok foglalkozni. Azt hiszem azonban, nem nagyon világtól elrugaszkodott az, ha abból indulok ki, hogy nem létezik olyan közgazdasági elmélet, amely a makroökonómiai döntéshozatalt „mérnöki problémává” változtatná, ahol az elvek tisztázottak, és pusztán csak a gyakorlati kivitelezésben lehetnek viták.⁴

A következőkben azt vizsgálom, hogyan járulhat hozzá a hivatalos statisztikai rendszer ahhoz, hogy az adatok valóban információt tartalmazzanak a gazdaságpolitikuskoknak és általában a makroközgazdászoknak. Felteszem az egyszerűen hangzó „Mit tegyen?” és a talán nem annyira népszerű „Mit ne tegyen?” kérdéseket. Azaz, tegyen-e nyilvánossá minél több alapadatot (nem törődve azzal, hogy az adat önmagában nem információ), és „méréskelje” az általános aggregált hivatalos statisztikák közlését (nemzeti számlák, GDP, CPI⁵)? Nyilván a második kérdésfeltevést kell részletesen indokolnom, ezért erről szól a cikkem túlnyomó része.

1.1. Az aggregációs paradigma

Egy az interneten terjedő szellemes mondás szerint: „A statisztikák olyanok, mint a bikini, amit felfednek, sokat ígérő, de az igazán lényeges az, amit elfednek.”⁶ Véleményem szerint ennek a tréfás formában kifejezett elégedetlenségnek az oka az, hogy mindmáig uralkodik a közgazdaságtanban és a gazdaságstatisztikában egy olyan empirikus paradigma, amit jobb híján aggregációs paradigmának fogok nevezni. Mit jelent ez pontosan?

A gazdasági statisztikák szinte kizárólag aggregátumok, alapadatokból képzett súlyozott átlagok, vagy összegek. (Összeg = átlag × tagok száma). Nem csak a statisztikai hivatalok aggregálnak, a különböző tőzsdeindexek sem különböznek jelleget tekintve az árindexektől vagy GDP-statisztikáktól. Az aggregáció paradigmájának nevezem azt a törekvést, amely arra irányul, hogy bizonyos fogalmakat egydimenziósan mérjünk, egy dimenzióba sűrített információval fejezzük ki. Ez egy természetes emberi törekvés, amelytől nem csupán a gazdaságban, de más területeken sem tudunk, és nem is kell feltétlenül szabadulni. A sokszor bonyolult formulák segítségével készült rangsorok számos sportban nyertek létjogosultságot, de léteznek egyetemi és kórházi rangsorok, korrupciós és demokrácia-indexek és számtalan más aggregátum, amelyet az erre szakosodott szervezetek rendszeresen újraszámolnak, és

⁴ Személyes tapasztalatom szerint ez túl optimista nézőpont a tulajdonképpeni mérnöki problémák szempontjából is, de itt csak a kontraszt az érdekes, és nem az, hogy pontosan mennyire fehér a fehér.

⁵ CPI (consumer price index): fogyasztói árindex.

⁶ Lásd http://thinkexist.com/quotation/statistics_are_like_a_bikini-that_they_reveal_is/222844.html.

amelyeket a sajtó előszeretettel közöl. Ezeknek az indexeknek, rangsoroknak számos funkciójuk lett az idők folyamán, de néhányat csak az újságok címlapján használnak. Hatásuk természetesen eldöntendő kérdés a konkrét index és terület esetében. Egy általános állításként azonban kimondhatjuk, hogy az általuk megragadni kívánt fogalom (kategória) nem annyira pontosan definiált, hogy az indexre egyértelmű „mérési utasítás” legyen adható.⁷ A gazdasági teljesítmény vagy az általános árszínvonal intuitíve jól értelmezhető fogalmak, ám – mint azt a gyakorlatban is látjuk – nincs két közgazdász, aki egyformán határozná meg ezeket, ha gyakorlati mérésre kerül sor. Bizonyos modellek bizonyos feltevések mellett egyértelmű definíciókat adnak ugyan, de a gyakorlatban senki sem gondolja komolyan, hogy ezek a feltételek teljesülnek.

Mindazonáltal az indexszámítók túlnyomó többsége filozófiai értelemben realistiként viszonyul a kérdéshez, azt hiszi, hogy objektíve létezik a gazdasági (vagy sport-) teljesítmény szintje, és keresi ennek minél pontosabb kifejezési formáját, mértékét. Ezért szokás arról beszélni, hogy egy adott mérték torzít vagy pontatlan, hiszen egyébként a pontatlanság, a mérési hiba nem lenne értelmes fogalom. Az aggregátumok alternatív felfogása inkább instrumentalista. E a felfogás szerint nem hiszünk az objektív teljesítmény létezésében, de azt gondoljuk, hogy az index (aggregátum) hasznos célokat szolgálhat. Például nem biztos, hogy avval kell törődnünk, hogy a tenis-világranglista pontszámai megmondják-e, ki a jobb teniszező és mennyivel, az a fontos, hogy a ranglista alapján történő kiemelések érdekesebb versenyekhez vezetnek. Az egyetemi rangsor segíthet tájékoztatni a hallgatókat arról, hogy milyen karrierkilátásaik lesznek, ha az egyik, vagy a másik egyetemre mennek, de nem szükséges azzal foglalkoznunk, hogy a ranglistán második egyetem valóban „objektíve” jobb-e, mint a 15. helyen álló.⁸ Azt gyanítom, hogy a közgazdasági aggregátumok készítői és felhasználói túlnyomórészt realisták, míg a sport- vagy egyéb indexek készítőiben több az instrumentalista beütés. A következőkben amellet fogok érvelni, hogy a filozófiai hozzáállás különbségei nem lényegtelenek, mert eltérő kérdésfeltevésekhez és megoldásokhoz vezetnek.

A közgazdasági aggregátumok esetében az is fontos, hogy bizonyos „hivatalos” statisztikai aggregátumoknak rendkívüli, hűsbavágó politikai fontossága lett az idők folyamán, mintegy intézményesültek. (A gazdaságstatisztika történetéről lásd *Hüttl* [2003].) Ilyenek például a hivatalos GDP- és CPI-számok, vagy az államadóságra vonatkozó adatok. Megjelent a nemzetek fölötti statisztikai hivatal is (az Eurostat), amelynek „mérési” nagy hatással lehetnek az odatartozó országok sorsára. A

⁷ Az ár- és volumenindex-számításnak nagy elméleti irodalma van, és a gyakorlati indexek szellemükben sokat felhasználnak ennek az eredményeiből, de nincs – mégpedig gyakorlati okokból – olyan létező közgazdasági indexszám, amely ezeknek pontosan eleget tenne.

⁸ A legtöbb közgazdasági aggregátum időben és objektumonként is aggregál. A sportokban használt aggregátumok gyakran csak időben, de ez a különbség itt most nem fontos.

Stiglitz-bizottság egyik fontos üzenete az volt az, hogy „amit mérsz, hat arra, amit csinálisz” (*Stiglitz–Sen–Fitoussi* [2009], *Stiglitz* [2015]). Az aggregátum azonban nem pusztán adat (amit mérünk), hanem tudással bővített adat, azaz információ, hiszen az, hogy mit adunk össze (például mit tekintünk jövedelemnek), vagy hogyan súlyozunk, már nem olvasható le a „mérőállásról”. Azonban, mint a statisztikai aggregátumok sorsa is bizonyítja, tudásunk nagyon is véges.⁹ A következő szakaszban ezt az állítást igyekszem alátámasztani.

1.2. Az aggregálás átka

Az aggregált GDP-t vagy annak növekedési ütemét információnak tekintjük, nem pusztán adatnak. Hatalmas erőforrásokat fordítanak arra, hogy ezeket a számokat létrehozzák, az alapadatok gyűjtésére és azok transzformálására. Kérdés azonban, hogy mire jók ezek a számok, van-e mögöttük igazi tudás, vagyis hasznosan tudjuk-e ezeket felhasználni? *P. A. Samuelson*nak tulajdonítják azt a költői kérdést, hogy mi történik a GDP-vel, ha egy professzor feleségül veszi a bejárónőjét? A laikust talán meglepi a válasz: a GDP csökken. Akármire is használjuk a GDP-t (a jólét mérése, a gazdaság ciklikus helyzetének megítélése, nemzetközi összehasonlítás, aggregált munkatermelékenység mérése stb.) a józan ész azt súgja, hogy ennek a magánszerződésnek nem kellene befolyásolnia elemzői vagy gazdaságpolitikai ítéletünket. Lerázhatjuk magunkról a problémát avval, hogy ez csak egy apró hiba – biztosan van ellenkező irányú hiba is – és az ügynek nincs olyan nagy jelentősége, hogy egy a társadalom iránt felelősséget érző egyetemi tanár emiatt megtartóztassa magát a frigyre lépéstől.

Ám legyünk machiavellisták, és keressünk manipulációs lehetőséget. Tegyük fel, hogy egy kormány egyik napról a másikra szeretné megnövelni az ország GDP-jét, és nem csak elhanyagolható mértékben. Felméri, hogy kik azok a nyugdíjasok, akik idősek klubjába járnak, vagy idősek otthonában élnek, ahol közösségi elfoglaltságaik vannak, beszélgetnek más nyugdíjasokkal stb., és a kormány ezektől az emberektől megvonja a nyugdíjat, egyidejűleg társadalmi munkásként alkalmazza őket abból a célból, hogy törődjenek más idős emberekkel, ha másként nem, legalább hallgassák meg őket.¹⁰ Fizetésüket úgy állapítják meg, hogy nettó jövedelmük ne változzék. Hazafias érzelmű nyugdíjasok szívesen belemennek ebbe a cserébe, hiszen szemé-

⁹ Egy lényeges különbség a fizikai mennyiségek (mint például a tömeg) mérése és a GDP mérése között: egy darab trappista sajt tömege is aggregátum, az őt alkotó molekulák tömegének aggregátuma. Azonban nem szükséges, és nem is praktikus, a molekulákat egyenként megmérni és a tömegeket összeadni. Az aggregált GDP viszont „megfigyelhetetlen” az összetevői megfigyelése nélkül. A példa egyébként mutatja, hogy valójában az adat sem teljesen „tudásmentes”. A trappista tömegét ugyan adatnak tekinthetjük, de maga a tömeg fogalma is szellemi vívmány, bármennyire is triviálisnak tűnik ma ez számunkra.

¹⁰ A pszichoanalitikusok is termelnek GDP-t, és egy főre jutóan talán nem is keveset.

lyes jövedelmük nem változik, viszont így hozzájárulhatnak a GDP növekedéséhez. Az ötlet megvalósítása nemcsak a GDP-t, hanem a foglalkoztatást is növelné, ami egy újabb jó pont. Sajnos a munkatermelékenységet inkább csökkentené, ami viszont rossz fényben tüntetné fel az országot azok szemében, akik a GDP/foglalkoztatás hányadost használják nemzetközi versenyképesség mérésére, de annyi baj legyen.

Lehet, hogy a példa utópisztikus, de a manipuláció lehetősége fennáll, és minden bizonnyal lehet kevésbé átlátszóan is csinálni.¹¹ A tudatos manipulációról még lesz szó, de a merőben véletlen jellegű, ám súlyos következményekkel látszó „torzításokra” számos példa van. Például a Stiglitz-bizottság jelentésében is hivatkozott *Hartwig* [2005] azt mutatja meg, hogy az 1990 és 2005 közti növekedési különbség az Egyesült Államok javára Európával szemben jelentős részben a különböző deflátor számítási filozófiáknak volt tulajdonítható. Ennek gazdaságpolitikai következményei is voltak, az Európai Unióban az Egyesült Államok mintaképpé vált, amelyet követni kívántak a gazdasági intézményrendszer reformja révén is. Van hozzánk közelebb eső példa is. *Hüttl–Surányi–Vita* [1998] cikkükben a közszolgáltatási szektor sajátos állapotának a hatását vizsgálják az átmenet után az egykori kelet-európai szocialista országok GDP-jének mérésére. Felhívják a figyelmet arra, hogy milyen nemkívánatos gazdaságpolitikai következményei (forráskivonás a közszolgáltatásokból) lehettek/voltak ennek.

A GDP az egyik leggyakrabban használt és rosszul használt mutató. A Stiglitz-bizottság megállapítja (*Stiglitz–Sen–Fitoussi* [2009]), hogy a GDP sem a gazdasági teljesítménynek, sem a jólétnek nem ideális mérőszáma, mégis gyakran ezekben a funkciókban használják. Rendkívül részletesen és pontosan összefoglalja a GDP felhasználásával kapcsolatos problémákat. Hangsúlyozza, hogy egy valamely célra kifejlesztett mérték más célokra való használata zavart okozhat. Leszögezi azt is, hogy nincs olyan indikátor, ami képes lenne egy olyan komplex struktúrát megragadni, mint a társadalom, és megállapítja, az a fontos, hogy mit akarunk mérni. A jelentés számos javaslatot tesz különféle mérési célokat (teljesítmény, jólét, fenntarthatóság) szem előtt tartva, és megjegyyezve, nincs egyértelmű szabály arra, hogy a gyakorlatban miként lehet ideális mércéket létrehozni, sőt maga az ideális mérce sem elérhető cél. Mégis úgy véli, hogy a fejlődés lehetséges, például a jólétet pontosabban mérő aggregátumok előállíthatók, vagyis a Bizottság is az aggregációs paradigma realista interpretációja talaján áll.

Ezen a ponton feltehetjük azt a kérdést, hogy milyen tudás rejlik a GDP statisztikai fogalmában. Úgy is kérdezhetnénk, hogy azok az elvek, amelyek alapján az újabb és újabb GDP-mérési standardok megszületnek, milyen elméletből táplálkoznak. A GDP mérése szorosan kötődik a ma egyre gyakrabban kárhozottatott neoklasz-

¹¹ Nem tudom, milyen alapja lenne, mondjuk az Eurostatnak arra, hogy ne vegye figyelembe a GDP-ben az időskorú szociális munkások teljesítményét, ha ezt az ötletet mégis megvalósítanák.

szikus általános egyensúlyelméleti paradigmához. A GDP alapkonceptiója szerint a GDP az egyéni jövedelmek összege, ahol a transzfereket igyekszünk megkülönböztetni a jövedelmektől, függetlenül a jövedelemszerző tevékenység jellegétől. Ebben a paradigmában mindenki vagy termel (és a munka valamely termelési függvény inputja), vagy fogyaszt. A rendőrség vagy az állami adminisztráció, az egészségügy vagy az oktatás ugyan beleerőltethető ezekbe a kategóriákba, de csak nagy erőfeszítéssel. Olyan ez, mintha a kémiában csak reagensekről és termékekről beszélnénk, és elfelejtkeznénk arról, hogy vannak katalizátorok is, amelyeknek az a feladata, hogy a reakciók simán lezajlódhassanak, és amelyekből van „optimális” mennyiség, ami sem nem túl sok, sem nem túl kevés. Kevesen gondolják ma már, hogy a neoklasszikus paradigma a közgazdasági gondolkodás megdönthetetlen csúcsa, de kevesen tudják azt, hogy amikor a makroökonómiai tényeket elemzik, akkor rabjai ennek a paradigmának.

Léteznek olyan vélemények, amelyek szerint számos probléma megoldódna, ha a mérés nem a GDP-re, hanem a vagyoni (wealth vagy capital) koncentrálna (*Bod* [2015]), *Stiglitz–Sen–Fitoussi* [2009]). A Stiglitz-bizottság egyik fontos állítása az, hogy a fenntartható növekedés egyszerű kritériuma az, hogy az egy főre jutó vagyon nem csökken. Vagyis szerintük a jól mért vagyon nagy információtartalommal rendelkező aggregátum. Azonban, ahogy maga Stiglitz is rámutat, a vagyon jelenlegi mérésének számos olyan buktatója van, amely ezt megakadályozza (a vállalati érték mérése buborékok estén, a járadékok „értékként” való elszámolása, illetve a recesszió idején történő emberi és egyéb tőke drasztikus csökkenésének nem megfigyelése). Ha jobban belegondolunk vagyoni fogalmunk ugyanolyan neoklasszikus, mint a GDP. Stiglitznek természetesen vannak ugyan ötletei a javításra, de egyértelmű helyes mérési utasítást nem tud adni, és – véleményem szerint – ilyen nem is létezik.

Eddig ugyan csak a jövedelem- és vagyonaggregátumokkal foglalkoztunk, de amit eddig leírtam megfelelő módosításokkal minden más aggregátumra is igaz. *Searle* [2015] például kimutatja, hogy a megélhetésiköltség-index, amitől Nagy-Britanniában a munkások millióinak bére függött, nem „pártatlan” statisztikai konstrukció volt, hanem politikai érdekek ütközésének a terméke. A közgazdászok alapvetően kétfajta módon viselkednek az aggregátumokkal kapcsolatban. Vagy elfogadják őket, mint adottságot, és igyekeznek nem is tudni, hogyan készülnek, vagy pedig veszik a fáradságot és igyekeznek a mélyükre nézni.¹² Ez utóbbi esetben elkerülhetetlen, hogy elvi és gyakorlati hibákat is találjanak, és lesznek ötleteik a tökéletesítésre. Csakhogy nem mindegy, hogy realista vagy instrumentalista módon állunk-e a kérdéshez. Azt hisszük-e, hogy van „elvben tökéletes” reáljövedelem-mérték, vagy

¹² A közgazdaságtan aránytalanul nagy része elméleti. *Zvi Griliches* [1986] egy családi történettel illusztrálta a közgazdasági adatok problémáit. „Nagyapám soha nem evett fasírtot. Étteremben azért nem, mert nem tudta mit raknak bele, otthon pedig azért nem, mert tudta.”

pedig úgy gondoljuk, hogy konkrét célkitűzéseink döntik le, hogy mit tekintünk reáljövedelemnek, amely csupán egy szakkifejezés, amelyet megszokásból használunk?

Véleményem szerint a realista megközelítés mindig újra fogja termelni a mértékek felhasználásával kapcsolatos problémákat. Az instrumentalista megközelítés nagy előnye, hogy konkrét problémákra, konkrét kérdésfeltevésekre koncentrálnak. Ez ugyan azt jelenti, hogy a kapott aggregátum gyakran egyedi (mintegy egyszer használatos) lesz. Vegyünk egy példát. Az árszint nemzetközi összehasonlításainak eredeti célja annak feltárása volt, hogy miként lehet olyan nominális árfolyamokat „bevezetni”, amelyek az egyes országok közös valutában mért árszintjét azonosítják. Szerintem ez egy „realista” cél. Egy pontos gyakorlati cél az lett volna, ha annak érdekében dolgoznak, hogy a nemzetközi arbitrázslehetőségeket minimalizáló árfolyamokat vezessenek be. Nem kell magyarázni, hogy ez a konkrét kérdésfeltevés határozott irányba terelné a gondolkodást. A vásárlóerőparitás-számítások mindmáig inkább a realista attitűdöt tükrözik, de amikor valódi gyakorlati problémát kell megoldani, akkor a józan ész az egyedi megoldások irányába hat. Ilyen gyakorlati megközelítésre ma is számos példa van. Például, amikor transznacionális vállalatok a különböző országokban dolgozó alkalmazottaik bérét igyekeznek meghatározni, tekintettel vannak az árkülönbségekre, de nem valamilyen absztrakt árszínvonal különbségére, hanem a konkrét alkalmazottak életét befolyásoló árak különbségére. Nem hiszem, hogy célul tűznék ki az ideális „vezérgazgató PPP”¹³ meghatározását, megelégednek azzal, hogy a bérezés ne okozzon konfliktusokat.

Eddig az aggregációs paradigma keretein belül maradtunk. A következő szakaszban arról lesz szó, hogy a hagyományos aggregálás mellett vannak más eljárások is nagy mennyiségű gazdasági adat értelmezéséhez.

1.3. Big data és az aggregációs paradigma

A gazdaságstatisztika létrejötte óta végbement informatikai fejlődésnek számos aspektusa közül a következők tűnnek témánk szempontjából leginkább relevánsnak. Először is lezajlott az adattárolás és -átvitel forradalma. A ténylegesen kezelhető adathalmazok méretének növekedése következtében a gazdasági adatok tárolása és kommunikációja szempontjából a lehetőségek szinte korlátlanok tekinthetők. A természettudományokban használt és használható adatbázisok méretéhez képest a nagy gazdasági adatbázisok eltörpülnek. Ezzel szemben a nemzeti statisztikai hivatalok kialakulásának idején az adatgyűjtés, -tárolás és -kommunikáció fizikai korlátai is olyan erősek voltak, hogy magánintézményektől nem volt elvárható ezek szolgáltatásaként nyújtása. Ma magáncégek (lásd Google Trends <https://www.google.hu/trends/>), sőt egyete-

¹³ PPP (purchasing power parity): vásárlóerő-paritás.

mek (az MIT¹⁴ „billion prices” projektje <http://bpp.mit.edu/>) is menedzselnek nyilvánosan hozzáférhető nagy gazdasági adatbázisokat.

A *big data* forradalma azonban nemcsak a nyersadatok beszerzésére, tárolására, átvitelére vonatkozik, hanem az adatok feldolgozására is. A hagyományos statisztikai paradigmák mellett számos új paradigma is megjelent (statistical learning theory, data mining címszavak alatt), illetve megvalósíthatóvá vált. (*James et al.* [2013], *Hastie–Tibshirani–Friedman* [2001]). Ez azt is jelenti, hogy ma az adatfeldolgozási módszereknek hatalmas tárháza áll rendelkezésre, amelyeket mind informatikai, mind pedig emberi humántőke-oldalról nézve a legkülönbözőbb intézmények képesek használni, és amelyek egymással versenyeznek anélkül, hogy kristálytiszta kirajzolódna melyik a legjobb, akárcsak egyes problémák megoldása szempontjából is.

Az előzők alapján talán természetesnek tűnik, hogy a statisztikai hivatalok tevékenységében egyre jelentősebb helyet foglaljon el a nagy tömegű alap- (nyers) adat szolgáltatása. A statisztikai hivatalok azonban továbbra is igyekeznek elsősorban feldolgozott adatokat szolgáltatni, amelyeknek – vélhetőleg – nagy az információ tartalma. Az egyedi adatok nyilvánosságra hozásának gátja a személyes adatok védelmének igénye, ami elsősorban politikai kérdés. Mégis, mi gátolja meg azt, hogy egy hivatal ne tegyen nyilvánossá minden alapadatot, betartva természetesen a törvényes korlátokat? Tovább mi indokolja azt, hogy a gazdasági adatok statisztikai hivatalok által közölt túlnyomó része (erősen) aggregált adat?¹⁵

Az újfajta adatelemzési módszerek, amelyeket gyakran statisztikai tanulási elméletnek szoktak nevezni, másfajta megközelítést is sugallnak. Alapvetően instrumentalista, gyakorlati problémák, konkrét kérdések megoldásában segítő számításokat végeznek, amelyben keletkezhetnek ugyan „aggregátumok”, azaz az adatokat tömörítő változók, de ezek nem önmagukban vett célok, és nem is szükséges eszközök. Ezeket az eljárásokat induktív módszereknek is nevezik, amelyek nem feltétlenül újak, de az informatika fejlődése, az adatok nagy tömegének kezelhetősége gyakorlatilag megvalósíthatóvá tette ezeket, ösztönözve a fejlődésüket (újabb és újabb módszerek létrejöttét), és egyszersmind be is bizonyította képességeiket. Az induktív gondolkodás ősrégi, de hagyományosan kevés esetből szoktunk általánosítani. Az induktív statisztikai módszerek akkor adnak újat, amikor nagy tömegű, nagy dimenziójú adathalmazokkal rendelkezünk, amelyeket az emberi agy a számítógép segítségével nélkül nem tud feldolgozni, nem képes belőlük általánosításokat megfogalmazni.

Hogyan működne az induktív megközelítés akkor, amikor a problémánk az, hogy adatokból próbálunk államháztartás-fenntarthatósági „kritériumot” találni, de ezt

¹⁴ MIT (Massachusetts Institute of Technology): Massachusettsi Műszaki Egyetem.

¹⁵ Nyilvánvalóan vannak fokozatok. A tej ára Magyarországon májusban is, mind időben, mind térben aggregált adat, ám nyilvánvalóan jóval kevésbé aggregált, mint a feldolgozatlan élelmiszerek árándexe.

nem egy predefiniált egy- (vagy akár kevés) dimenziós mutató (lásd államadóság/GDP) útján próbáljuk megtenni? Például összegyűjtünk olyan ország eseteket a múltból, ahol voltak adósságválságok és olyanokat, ahol nem. Természetesen az utóbbira több a példa, de elég, ha az előzőre is vannak példák. A teljes minta elemszáma legyen N . Az adott esetekhez hozzárendeljük az országokról rendelkezésre álló relevánsnak tűnő információt nagy dezaggregáltságban, azaz minden esethez egy p dimenziós megfigyelési vektor tartozik. Valamilyen jól bevált klasszifikációs algoritlussal (*Hastie–Tibshirani–Friedman* [2001]) a rendelkezésre álló minta egy részhalmazán végrehajtjuk a klasszifikációt, azaz az adatok alapján létrehozunk egy leképezést, ami minden lehetséges ismérvekvektorhoz hozzárendeli a 0 („válság”) vagy 1 („normális állapot”) értékeket. Ezután a minta fennmaradó részén teszteli a modellt, vagyis ellenőrzi, hogy a „becsült” klasszifikáció milyen mértékben jelzi előre (természetesen visszamenőleg) a becslés során nem használt adatokat. Többfajta modellt is becsülhetünk és tesztelhetünk, majd valamilyen kritérium alapján a legjobb predikciós képességűt választjuk. Ez a módszer nem feltétlenül hoz létre „aggregátumot”, azaz nincs szükségképpen olyan egydimenziós mennyiség, amely összesűríti az adatokat, és „kifejezi” a fenntarthatóságot. Ez tehát megkülönbözteti az aggregációs paradigmától.

A klasszifikációs probléma persze nem új, régóta léteznek olyan statisztikai-ökonometriai modellek, amelyek ennek megoldására születtek. Az új módszerek azonban gyakran matematikai struktúrájukat tekintve is alapvetően különböznek a régebbiektől (például nincs explicit valószínűség-számítási hipotézis mögöttük). Másfelől a metodológia szerves részét alkotja a modellválasztási procedúra, amely a hagyományos ökonometriában inkább mellékes szerepet kapott. Harmadrészt itt a hangsúly alapvetően a predikción van, és egyes paraméterek identifikálása másodlagos, míg a hagyományos ökonometriában ez fordítva van. A statisztikai tanulási elmélet központi „filozófiai” kategóriája a torzítás-variancia tradeoff,¹⁶ amely az ökonometriai irodalomban fehér hollóként jelenik meg.

Konkrét problémánk szempontjából azonban talán a legfontosabb az, hogy ezek a módszerek kezelik a $p > N$ eseteket is (az úgynevezett regularizációs elv alapján), azaz a változók száma meghaladhatja a megfigyelések számát. Ezzel szemben a hagyományos statisztikai módszerek szinte megkövetelik az aggregálást, hiszen gyakran csak így lehet a priori redukálni a változók számát úgy, hogy $p \ll N$ fennálljon, azaz sokkal több megfigyelésünk legyen, mint változónk.

Gyakran szokták ezeket az újabb módszereket tisztán adatvezéreltnek nevezni, de ez természetesen csak reklám. Tisztán adatvezérelt, minden a priori hipotézis nélküli modellezés, nem létezik, de annyi igaz, hogy kevesebb elméleti feltevés kell az alkalmazásukhoz. Nagy előny, hiszen a közgazdaságtanban nincs kísérletekkel meg-

¹⁶ Ha tudjuk is azt, hogy a becslésünk torzít, (amennyiben csak kicsit szór), akkor ez bizonyos célok szempontjából előnyös lehet. (*Hastie–Tibshirani–Friedman* [2001]).

alapozott kvantitatív tudásunk. Aggódni azonban nem kell, a közgazdasági problémák vizsgálata nem csak informatikai, statisztikai szaktudást kíván, a sikeres statisztikai modellekhez mindig szükség lesz tárgyismeretre is.

Az indukciós módszerekkel kapcsolatban eddig számos ellenérvés volt a közgazdászokban, ám jóval kevesebb a rokon gazdálkodástudományi területeken. Egyrészt sokfajta módszer van, és nem világos, hogy egy adott problémára mi a legalkalmasabb. Ez azonban a gyakorlatban kialakulhat. Másrészt kifogásolják, hogy az eredmények nem interpretálhatók, és különösen hiányolják az oksági összefüggések felderítésének hiányát. Ez utóbbi azonban nem feltétlenül reménytelen, erről lásd *Varian* [2014] ötleteit. Itt szerintem olyan gondolkodásbeli berögződésekről van szó, amelyeket nem árt felülvizsgálni. Az interpretálhatóság jelentése gyakran az, hogy tudunk-e értelmezni valamely eredményt a hagyományos elméleti keretekben, és nem biztos, hogy erre szükség van. A makroökonómiában az okság kérdéseinek tisztázásához nem csekély mértékben hozzájárulhatnak az induktív módszerek.¹⁷

1.4. Kitérő: a görög példa

Az aggregálás átkára jó példa Görögország és az EU kapcsolata. Természetesen nem akarom azt állítani, hogy a „görög tragédia” az aggregációs elvnek és annak a következménye, hogy a hivatalos aggregátumoknak intézményesült szerepe van, de azt igen, hogy az eset példája az aggregátumokon alapuló uniós szabályozás kudarcának. Vagyis azt mondom, hogy amennyiben azt a tanulságot vonnák le a Görögországgal kapcsolatos kudarcokból, hogy jobb mutatókra van szükség, akkor ezzel nem előznék meg a hasonló jövőbeli bonyodalmakat.

Az államadósság/GDP-hányados lényegében a költségvetési politika „fenntarthatóságának” megragadására született. Az EU intézményesített mechanizmusaiban (például maastrichti kritériumok, stabilizációs paktum) ez az aggregátum olyan fontos szerepet kap, amely minden országot érdekeltté tesz abban, hogy „jó színben” tüntesse fel.

Hogyan működött ez a mutató Görögország esetében? Az államadóssági és költségvetési adatokat a görög kormány (statisztikai hivatal) manipulálta, mégpedig abban az értelemben, hogy nem tartotta be az elszámolási szabályokat, és ez csak önbevallás útján derült ki (*Nechio* [2010]). A nyílt manipuláció mellett az adósság csökkentésére természetesen vannak kifinomultabb eszközök is (megfelelő pénzügyi szerződések formájában), amelyeket nemcsak államok, de vállalatok is előszeretettel alkalmaznak. Görögország kapcsán a kilencvenes években is felmerült ez,

¹⁷ Leegyszerűsítve azok számára, aki nem lennének benne járatosak: a modern makroökonómiában az alapvető „okok” exogén sztochasztikus folyamatok (népszerű nevükön „sokkok”), amelyek nem megfigyelhetők, és teljesen függetlenek az adott gazdaság működésétől.

felvetve a beruházási bankok felelősségét, amelyek – nem ingyen – de segítették abban, hogy az euró bevezethetősége érdekében jobb adósságmutatóval rendelkezessék.¹⁸ Az adósság általában – és nem csak az államadósság – olyan fogalom, amelyet szeretünk ugyan egy dimenzióban kifejezni, de mindenki tudja, hogy az egyetlen dimenzió nem tartalmaz minden fontos információt. De míg egy vállalatot elemző befektetőt semmi sem kötelez arra, hogy ne próbáljon meg a vállalati adósság struktúrájába betekinteni, addig a hivatalos államadósság-aggregátum szinte megfellebbezhetetlen azokban a kontextusokban, ahol annak szerződésekben rögzített funkciója van.¹⁹

A hányadosnak persze van nevezője is. Az 1995 utáni jelentős görög GDP-növekedés idején a hányados is elfogadhatóan alakult,²⁰ míg a 2009-es recesszió után nagyon rosszul, ami a görög állami pénzügyek fenntarthatóságát fokozottan rossz színben tüntette fel. Vajon miért jó a fenntarthatóság mérése szempontjából az adósságot a GDP-vel osztani?

Az első közvetlenül adódó válasz az, hogy a GDP a teljesítmény mérőszáma. Már eddig is láttuk, hogy milyen kételyek merülhetnek fel evvel kapcsolatban, de Görögország esetében van egy eddig nem tárgyalt probléma is. Úgy becsülik, hogy az adott GDP-definíció mellett a görög GDP több, mint 20 százaléka az „árnyékgazdaság” (*Schneider* [2013])²¹. Az ennek nem elhanyagolható részét kitevő feketegazdaság mérésére szolgáló elvek is elég bizonytalanok,²² de az adatok szükségképpen még inkább.²³ Ekkora bizonytalanság mellett még a legbecsületesebb statisztikai hivatal is „félremérhet”, és mi okunk van feltételezni, hogy a statisztikai hivatalok kevésbé vannak politikai befolyás alatt, mint a központi bankok, bíróságok stb.?

Van egy másik alapvető kérdés is: miért gondoljuk, hogy a görög kormány rendelkezik a teljes belső jövedelem felett? Egy vállalat esetében az adósság/nettó cash flow-hányados azért elfogadható adósság-fenntarthatósági mutató, mert többkevesebb joggal tételezhető fel, hogy egy vállalat azt tesz a nettó cash flow-val, amit akar. Viszont egy olyan állam, ahol ekkora a feketegazdaság, és ahol közismerten nehéz adókat behajtani, bizonyára nem diszponál a teljes jövedelem felett.

¹⁸ Lásd <http://www.theguardian.com/business/2010/feb/28/goldman-sachs-investment-banking-greece>.

¹⁹ Az államadósság egyik legvitatottabb kérdése a nyugdíjrendszerekben rejlő (implicit) államadósság. Ennek hivatalosan adósságként való elfogadása vagy nem elfogadása teljesen indokolatlanul ösztönöz bizonyos gazdaságpolitikai beavatkozásokra. Lásd a magyar magán-nyugdíjpénztárak esetét.

²⁰ A Nemzetközi Valutaalpnál beható elemzéssel igyekeztek kizárni a növekedés átmeneti tényezőit és végül is arra a következtetésre jutottak, hogy a görög potenciális növekedési ütem a tekintélyt érdemlő éves 3,5 százalékra nőtt (*IMF* [2006]).

²¹ De léteznek becslések, amelyek 36 százalékra teszik, lásd http://ec.europa.eu/europe2020/pdf/themes/07_shadow_economy.pdf

²² Például a védelmi pénzt szedő bűnöző tevékenysége miatt nem szolgáltatás, miként a rendőre is? Ha a Keresztapa 2-nek lehet hinni, akkor ilyen esetben néha a közszolgáltatások hiányosságainak pótlásáról van szó.

²³ Erről Magyarország esetében lásd az itt látható linkben idézett „tényfeltáró” cikket: http://hvg.hu/gazdasag/20121109_Szex_drog_es_GDP_itt_az_osszefugges

Ha a GDP helyett egy másik, „a kormány számára rendelkezésre álló jövedelem” szerepelt volna nevezőben, akkor talán már az adósságválság előtti években is más-ként kezelték volna Görögországot. Könnyű felhívni a figyelmet arra, hogy egy má-sik mutató ebben az esetben jobb lett volna, de mi is pontosan ez a mutató, hogyan határoznánk meg, és ha kiderül, hogy annak is vannak hátulütői – és ez szükségszerű – hogyan módosítanánk? Azt gondolom, hogy ezekre a problémákra nincs jó köz-pontosított megoldás. Az egyes piaci és nem piaci szereplők persze definiálhatnak fenntarthatósági mutatókat, és ezeket folytonosan frissíthetik, de viszonyulhatnak a problémához úgy is, ahogyan az előző fejezetben láttuk, vagyis mutató nélkül, köz-vetlenül is alkothatnak ítéleteket az adatok alapján arról, hogy az államadósság fenntartható-e. Előfordulhat persze az is, hogy bizonyos megoldásokat sokan adaptálnak, és azok mintegy standarddá válnak, de a központosított megoldás szükségképpen rugalmatlansággal jár.

2. Következtetés

Tekintsünk két tényállítást: 1. Kovács Botond 2014. 08. 30-án két üveg Borsodi vi-lágos sört vásárolt 156 forint/üveg áron, az X úton található Y üzletben. 2. Magyaror-szágon a lakossági sörfogyasztás 2014-ben, folyó áron, Z milliárd forint volt. A makroközgazdászoknak az 1. állítás pusztán adat, amivel nem tudnak semmit sem kez-deni. A 2. állítás azonban információ (informatív adat), amit beépítenek elemzéseikbe, modelljeikbe, előrejelzéseikbe. A 2. állítást a statisztikai rendszerek generálják, input-ként az 1. állításhoz hasonló egyszerű tények sokaságát használják fel, de még sokkal-sokkal több mindent. Egy egész fogalmi rendszer alapján működnek, amelyet az éppen érvényes módszertani útmutatók váltanak aprópénzre. De még ez sem igazán aprópénz, a legrészletesebb metodológiai útmutatóból sem lehet rekonstruálni egyértelműen, hogyan állna elő a 2. állítás. A két állítás között persze sok más is elhelyezkedhet. Péld-ául: 3. Magyarországon 2014. 3. negyedévében a sörfogyasztás W liter volt. Ez az adat szintén aggregátum, amely sok kis egyszerű 1. típusú állításból következik, de mégis jelentősen eltér attól. A mögöttes módszertan ugyanis sokkal egyszerűbb, a mé-rési utasítás majdnem egyértelműen értelmezhető. (Nyilvánvalóan definiálni kell a sör fogalmát, és fel kell tételeznünk, hogy a sört nem tartogatjuk, hanem viszonylag gyorsan megisszuk a vásárlás után.) Ma a statisztikai adatszolgáltatás túlnyomó részben a 2. típusú adatokra koncentrál, és ennek jó okai vannak. A felhasználók számára ezek informatívak, és az adattárolási és feldolgozási munka jelentős részét a statisztikai hivatal vállalja magára. Lehet, hogy ez a 3. típusú adat nem informatív ma egyetlen makroközgazdász számára sem, mégis azt állítom, hogy jó esély van rá, hogy a big

data-forradalom eredményeképpen a 3. állítás és a hozzá hasonló tények belátható időn belül informatívak lesznek, míg az 1. típusú állítások, érzésem szerint, még hosszú ideig csak az *Y* üzlet marketingesei számára jelentenek információt.

Ez tehát az első szakaszban említett pozitív állításom lényege. Egyre több, eddig nem informatív adat válik (potenciálisan) informatívá, ezért indokolt az, hogy a statisztikai hivatalok adatgyűjtő és -feldolgozó tevékenységük inputjait és outputjait egyre részletesebben nyilvánossá tegyék, ha nem is kell feltétlenül a „végső” alapadatokig elmenni. Fontos, hogy az adat informatív tétele munkát igényel. Megtörténhet, hogy először egy adattal nem tudunk mit kezdeni, de ha dolgozunk a problémán, akkor eredményre juthatunk. Ezért nem jó az a kifogás, hogy fölösleges magas dezaggregáltságú adatokat közölni, mert a felhasználók úgysem tudnak mit kezdeni vele.

Ugyanakkor van egy másik állításom is, nevezetesen az, hogy az olyan „magas szintű” aggregátumokra, mint a nemzeti számla aggregátumok, vagy a különböző általános árindexek (vagyis a 2. típusú adatok) kevesebb figyelmet kellene szentelni, sőt, ha radikálisan akarom megfogalmazni a véleményemet, akkor szerintem akár hosszú távon el is lehetne őket hagyni.

Az előzőkben részletesen foglalkoztam az aggregálás átkával, és igyekeztem bizonyítani, hogy a közgazdasági aggregátumok használata számos nem kívánatos következménnyel jár, különösen, ha azok „szentesített”, hivatalos statisztikák. Amellett is érveltem, hogy ezek „javíthatatlanok” abban az értelemben, hogy esetükben nincs sok értelme fejlődésről beszélni, mert maga az ideális aggregátum fogalma sem értelmezhető. Vigyázat, soha nem állítottam, hogy az aggregálás, adatsűrítés, indexszámolás értelmetlen! Azt mondtam, hogy ezek értelme a konkrét problémától függ, és ezért az indexek vagy aggregátumok leginkább egyszer használatos, vagy folyamatosan igazodó számítások eredményei lesznek. A statisztikai hivatalok többnyire nem alkalmasak az ilyen feladat elvégzésére, kivéve, ha a hivatalnak magának használnia is kell az aggregátumokat, konkrét problémák megoldására, nem csupán közölnie azokat. Ez lehetséges. Például elképzelhető lehet, hogy a statisztikai hivatal feladata annak eldöntése, hogy bizonyos hibahatáron belül a költségvetési terv realisztikus-e. Könnyen lehet, hogy ekkor a hivatal használ gazdasági aggregátumokat, de előbb-utóbb rájön, hogy az aggregálás elveit nem elsősorban a priori kell meghatározni, hanem az elvégzendő feladat függvényében.

A magas szintű aggregálás leszámolás mellett szól az is, hogy az adatelemző módszerek egyre kevésbé igénylik az aggregálást, az a priori adatsűrítés egyre kevésbé lesz szükséges. Az aggregáció megkerülésének igénye egyébként már ma is jelen van a gazdaságpolitikában. Úgy tűnik polgárjogot kapott például az a felismerés, hogy egy olyan komplex rendszernek, mint a bankrendszer, nem lehet egyetlen számmal kifejezni az egészségi állapotát. A ma már széles körben használt stresszteszt megkerüli az aggregálást, mintegy közvetlenül igyekeznek ítéletet mondani egy konkrét kérdésfeltevés esetében.

Mit kezdünk azzal az ellenvetéssel, hogy „de akkor mi lesz a maastrichti kritériumokkal, a stabilitási paktummal” stb.? Ennek a problémának a megoldása nem tartozik ide, de azt hiszem az uniós szerződésekben foglalt kvantitatív makrogazdasági kritériumok eleve pótlékként jöttek létre. A közös fiskális politika és a közös kultúra (például hogyan viszonyulunk az adatok manipulációjához) pótlékként. A pótlék csinosíthatása helyett jobb lenne ezeket a lényegesebb kérdéseket megoldani, ha lehet. Ha nem, akkor a pótlék sem fog segíteni.

Még egyszer összefoglalva: a statisztikai hivatalok eddig sok adatot gyűjtöttek és ehhez képest keveset publikáltak, viszont sokat számoltak velük. Több adatot kellene gyűjteniük és publikálniuk, de kevesebbet számolniuk. Ez lenne az alapfeladatuk, ami inkább informatikai, mint közgazdasági tevékenység. Ez nem zárja ki, hogy más feladatokat is ellássanak, és továbbra is hozzájáruljanak a gazdaságpolitikai döntésekhez magas információtartalmú adatok közlésével is, ám ezek az eddigiekhez képest más típusú erőfeszítéseket igényelnének.

Irodalom

- AYERS, I. [2007]: *Super Crunchers*. Bantam. New York.
- BOD P. Á. [2015]: Többet, mást, hamarabb – amit a válság utáni gazdaságpolitikai döntéshozatal igényel a statisztikától. *Statisztikai Szemle*. 93. évf. 7. sz. 713–724. old.
- GRILICHES, Z. [1986]: Economic Data Issues. In: *Griliches, Z. – Intriligator, M. D. (eds.): Handbook of Econometrics*. Vol. 3. Elsevier Science. Amsterdam. pp. 1465–1514.
- HARTWIG, J. [2005]: *On Misusing National Accounts Data for Governance Purposes*. (No. 101.). Arbeitspapiere. Konjunkturforschungsstelle, Eidgenössische Technische Hochschule. Zürich.
- HASTIE, T. – TIBSHIRANI, R. – FRIEDMAN, J. [2001]: *The Elements of Statistical Learning*. Springer. New York.
- HÜTTL A. [2003]: A gazdasági mérés történetéről. Adatok, elmélet, gazdaságpolitika. *Közgazdasági Szemle*. 50. évf. 2. sz. 164–182. old.
- HÜTTL A. – SURÁNYI B. – VITA L. [1998]: A gazdasági növekedés és fejlettség paradoxona a kelet-közép-európai átmenetben. *Közgazdasági Szemle*. VL. évf. 11. sz. 973–988. old.
- IMF (INTERNATIONAL MONETARY FUND) [2006]: *Greece: Selected Issues. Country Report*. No.06/5. Washington, D.C. <https://www.imf.org/external/pubs/ft/scr/2006/cr0605.pdf>
- JAMES, G. – WITTEN, D. – HASTIE, T. – TIBSHIRANI, R. [2013]: *An Introduction to Statistical Learning*. Springer. New York.
- NECHIO, F. [2010]: *The Greek Crisis: Argentina revisited*. FRBSF Economic Letter. No. 33. <http://www.frbsf.org/economic-research/publications/economic-letter/2010/november/greece-argentina-crisis/>
- SCHNEIDER, F. [2013]: *Size and Development of the Shadow Economy of 31 European and 5 Other OECD Countries from 2003 to 2013: A Further Decline*. Johannes Kepler Universität. Linz.
- SEARLE, R. [2015]: Is There Anything Real About Real Wages? A History of the Official British cost of Living Index, 1914–62. *The Economic History Review*. Vol. 68. No. 1. pp. 145–166.

- STIGLITZ, J. E. [2015]: *The Measurement of Wealth: Recessions, Sustainability and Inequality*. NBER Working Paper. No. 21327. Cambridge. <http://www.nber.org/papers/w21327.pdf>
- STIGLITZ, J. – SEN, A. – FITOUSSI, J. P. [2009]: *The Measurement of Economic Performance and Social Progress Revisited. Reflections and Overview*. Commission on the Measurement of Economic Performance and Social Progress. Paris.
- VARIAN, H. R. [2014]: Big Data: New Tricks for Econometrics. *The Journal of Economic Perspectives*. Vol. 28. No. 2. pp. 3–27.