

Regressziós modellek becslése és tesztelése Excel-parancsfájl segítségével (szoftverismertetés)*

Kehl Dániel,
a Pécsi Tudományegyetem
egyetemi tanársegéde
E-mail: kehd@ktk.pte.hu

Dr. Sipos Béla,
a Pécsi Tudományegyetem
egyetemi tanára
E-mail: sipos@ktk.pte.hu

A regressziószámítás az egyik legegyszerűbb és leggyakrabban alkalmazott ökonometriai módszer. A szerzők a lineáris regressziós analízis módszeréhez kidolgoztak egy Excel-környezetű parancsfájl, amelynek részletes használati-értelemezési útmutatója e cikk. Az Excel-parancsfájl felhasználási lehetőségét példák is illusztrálják. A programcsomag a statisztikai elemzés és modellezés graduális oktatásának hasznos eszköze lehet.

A parancsfájl letölthető a PTE-KTK honlapjáról (Excel-parancsfájlok felhasználása statisztikai elemzésekhez, kézikönyv és a BSC.zip és MSC.zip Excel-parancsfájlok, szám szerint 35, köztük az ismertetésre kerülő regresszio.xls).¹

TÁRGYSZÓ:
Regressziószámítás.
Statisztikai módszertan.
Excel.

* A tanulmányban előforduló esetleges hibákért kizárólag a szerzőket terheli felelősség.

¹ <http://www.gmi.ktk.pte.hu/index.php?mid=33#SiposB>.

A valós méretű statisztikai modellek, ezen belül a többváltozós regressziós feladatok megoldása kézi számításokkal általában nem, vagy csak nehezen végezhető el. A számítógépes feldolgozás lehetősége azonban új utakat nyitott meg a statisztika tudományában is. Napjainkban a számolási igény – a személyi számítógépek megjelenése és elterjedése miatt – már nem jelent különösebb akadályt, a számítások megkönnyítésére matematikai-statisztikai és ökonometriai szoftverek léteznek.

A jelenleg legnépszerűbb irodai programcsomag, a Microsoft Office változata 1990-ben jelent meg. A Microsoft Office (*Baczoni* [2007], *Bártfai* [2002]) és ezen belül az MS Excel (továbbiakban Excel) világviszonylatban és Magyarországon is széleskörűen alkalmazott szoftver. E program sok statisztikai műveletet képes elvégezni, és az alapfunkciókon túl, függvények segítségével felépíthetők a bonyolultabb statisztikai és ökonometriai módszerek is. További előny, hogy a módszerek, a felhasznált képletek alakíthatók, az adott feladat megoldásához testre szabhatók, láthatóvá és követhetővé válnak a részeredmények és a mellékszámítások. Az Excel – a speciális statisztikai szoftverekhez hasonlóan, de messze nem olyan részletességgel – a statisztika módszertanának nagy részét felöleli beépített modulja (Analysis ToolPak) segítségével, de több apróbb hiba (például rossz vagy félreérthető magyarra fordítás) és hiányosság is a sajátja. Az említett félrefordításoknál nagyobb hibák is megfigyelhetők, melyek a program korábbi verzióiban csakúgy megtalálhatók, mint a legújabbakban: a következtetési statisztikában oly fontos eloszlások esetén némely speciális esetben hibás, félrevezető értékeket szolgáltat. A témakör bőséges irodalommal rendelkezik, tanulmányunkban csak utalunk *Knüsel* ([1998], [2002], [2005]), *McCullough* és *Wilson* ([1999], [2002]) és *Yalta* [2008] munkáira, melyekből kimerítő „hibalista” meríthető. Az említett, több éve ismert hiányosságokat a jelentős tudományos kritika ellenére sem javították még ki. Ugyanakkor az alkalmazás kétségtelen és messze legfontosabb előnye, hogy az Office-csomag elterjedése miatt szinte mindenhol megtalálható.

Megemlítjük továbbá azt a fontos ténytet, hogy a statisztika oktatásában Magyarországon az egyetemeken és főiskolákon az Excel, mint táblázatkezelő szoftver elterjedt, főként könnyű elérhetősége okán (lásd e témában *Rappai* [2001]. Ismereteink szerint csak az Excel alapszolgáltatásainak használata terjedt el az oktatásban és az üzleti életben Magyarországon (*Balázsne Mócsai–Csetényi* [2003], *Jánosa* [2005]), pedig a program ennél többre képes, lehet batch file-okat, kötegelt parancsállományokat (a továbbiakban parancsfájlokat, illetve programokat) készíteni.

Tanulmányunkkal kapcsolódni kívánunk a *Rappai Gábor* által indított szakmai beszélgetéshez, ami a statisztikaoktatás átalakulásával, átalakításával foglalkozik. Az informatikai támogatottsággal és az Excel felhasználásával kapcsolatban Rappai a

következőket mondja: „meggyőződésem szerint a legszélesebb körben rendelkezésre álló támogatóeszköz használata a legindokoltabb” (Rappai [2008] 840. old.). A modernizáció jelentőségére hívja fel a figyelmet Kovács Péter [2008b] tanulmánya is, aki a Szegedi Tudományegyetemen bevezetett tanterven keresztül mutatja be a szegedi modellt, ami szintén erősen támaszkodik az Excelre. Úgy gondoljuk, hogy az általunk felvázolt Excel-alapú oktatás – melynek az egyik szelete a bemutatandó parancsfájl – az egyik, természetesen nem kizárólagos irány lehet a jövőben. A szakmai közösség tagjait továbbra is bízattjuk tapasztalataik és javaslataik megtételére.

Rátérve az alkalmazási lehetőségekre, véleményünk szerint az adatelemzés öt szintje oldható meg az Excellel.

Az *első szint* az, amikor a Függvény beszúrása varázslót (ikont) használjuk, tehát beépített statisztikai, matematikai és trigonometriai, mátrix, adatbázis stb. függvényeket alkalmazunk. A *második szint*, amikor az Eszközök/Adatelemzés² menüpont szolgáltatásait (például korrelációanalízis, regresszió stb.) használjuk. A *harmadik szint*, amikor magunk írunk konkrét adatsorhoz vagy adatsorokhoz képleteket, mivel nem minden feladathoz áll rendelkezésre beépített függvény. A *negyedik szint* az, amikor parancsfájlokat készítünk – vagyis a harmadik szintet általánosítjuk –, amelyek segítségével az általunk megadott adatbázis terjedelméig új adatbázisok felhasználásával korlátlan számban végezhetünk számításokat a programozott képletekkel, illetve függvényekkel. Gyakran igen sok számítást kell elvégezni. Eben az esetben az idővel való takarékos gazdálkodás a cél, mert a harmadik szintnél egy feladatsor számításainak elvégzése sokszor több óra vagy nap is lehet, amit a parancsfájlok felhasználásával egy perc alatt el lehet végezni. Az *ötödik szint* az, amikor a feladat a hagyományos módon nem oldható meg. Erre példa a CES (constant elasticity of substitution – konstans helyettesítési rugalmasságú függvény) termelési függvény, ahol a változók száma több mint a rendelkezésre álló egyenletek száma. A feladat a legjobban illeszkedő függvény paramétereinek megkeresése.³ A logisztikus és egyéb speciális trendfüggvények esetében a függvényeket nem lehet lineárisra transzformálni, a cél megkeresni azokat a paramétereket, amelyek mellett az illesztés a legpontosabb.⁴ A logisztikus regressziós függvények sem linearizálhatók, de iterációs eljárással a paraméterek becsülhetők, meghatározható olyan függvény, ahol a többszörös determinációs együttható elégségesen nagy. Az Excel a Visual Basic for Applications (VBA) felhasználásával programozható, így ezek a feladatok megoldhatók.

A szoftverek alkalmazásának egyik legnagyobb problémáját abban látjuk, hogy a számítási lépések nem követhetők, a felhasználó nem minden esetben érzékeli, hogy az adatok és azok kismértékű változásai hogyan hatnak az eredményre. Az általunk

² Az Eszközök/Bővítménykezelő/Data Analysis Toolpak hozzáadása után.

³ Lásd a CES-függvény becslését, ha három normálegyenlet áll rendelkezésre és a becsült paraméterek száma öt. A parancsfájl: ces1.xls

⁴ Kehl-Sipos [2009] és logisztikusregresszio.xls.

elkészített parancsfájlok⁵ – véleményünk szerint – kiküszöbölik ezt a hiányosságot: egyetlen cella, vagy vezérlőelem (Checkbox, legördülő menü stb.) módosításakor nyomon követhetjük az eredmények változását.

A munkalapokat egységes szerkezetben építettük fel. A változtatható, illetve megadható vagy megadandó adatokat sárga mezők jelölik, az eredményeket pedig egységes struktúrában jelenítettük meg. A végeredmények és az egyes cellák számításához használt képletek valamennyi esetben láthatók.

1. A regressziós Excel-parancsfájl működésének bemutatása

A regressziós modell készítésének (Hajdu *et al.* [1994–1995] 110–111. old.) első lépése a specifikáció, ami alatt a jelenséget leíró, modellben szereplő eredmény- és magyarázóváltozók kiválasztását, valamint a függvény konkrét formájának meghatározását értjük. Fontos szerepet játszik a specifikáció szakaszában az adatbázis, amelynek minősége, szerkezete nagymértékben befolyásolja a folyamat eredményességét. A gyakorlati munkában idősoros és keresztmetszeti adatokkal dolgozhatunk, ennek a modell feltételrendszerének ellenőrzésekor lesz jelentősége. Panel adatbázisokkal jelen anyagunkban nem foglalkozunk.

A specifikáció munkafázisának lezárása után a számításokat a regresszio.xls parancsfájllal lehet elvégezni. Ennek fontosabb lépései a következők:

1. A regressziós paraméterek becslése a klasszikus legkisebb négyzetek módszerével, melynek feltételei:

- a) a magyarázóváltozók nem sztochasztikusak, tehát mérési hibát nem tartalmaznak és lineárisan függetlenek (multikollinearitás hiánya),
- b) a hibatényezők (hibatagok, reziduumok) várható értéke 0, varianciájuk konstans, normális eloszlásúak és nem autokorreláltak.

2. A modell feltételrendszerének ellenőrzése. Ez a munkafázis visszahat mind a specifikációra, mind a paraméterbecslésre. Ebben a munkaszakaszban a modellező megállapítja, hogy adott szignifikanciaszint mellett mennyire fogadható el a modell. A fontosabb hipotézisellenőrzések: a regressziós modell paramétereinek globális és parciális tesztelése (a paraméterbecslés pontosságának vizsgálata, a paraméterek standard hibája, konfidencia intervalluma stb.), valamint a reziduumok vizsgálata: az autokorreláció és a homoszkedaszticitás tesztje, a ma-

⁵ Meg kívánjuk jegyezni, hogy elsősorban oktatási célból, de a gyakorlati alkalmazásokat is segitendő, 36 Excel parancsfájlt dolgoztunk ki, ezek egy része az alapképzésben használható fel, más része az ökonometriai jellegű tárgyokban használható. Az internetes hozzáférést biztosítottuk.

gyarázóváltozók közötti kapcsolat szorossága, a multikollinearitás ellenőrzése. A próbákkal nyert információk alapján döntést lehet hozni a modell esetleges megváltoztatásáról vagy a becslési módszer módosításáról. Ezek a döntések természetesen visszahatnak a specifikációra és indokolt esetben az egész eljárás (specifikáció, becslés, hipotézisellenőrzés) megismétlését igényelhetik.

3. A regressziós modell felhasználása elemzésre és előrejelzésre.

4. A verifikálás, aminek során a modellt szembesítjük a valósággal.

A program a bemutatásra kerülő regressziószámítást maximum 16 magyarázóváltozó és 2000 megfigyelés esetében végzi el.⁶ A programban a munkalapokon megjelenő színeknek jelentése van. A halványsárga cellák változtathatók, itt történik meg az adatok bevitele, a kívánt szignifikanciaszint beállítása, valamint a becslés/előrejelzés alapadatainak megadása. A tesztek végeredményei színes számokkal jelennek meg a fájlban. A modell ellenőrzésénél háromféle szint alkalmaztunk, a zavaró eredmények piros, a megfelelők zöld, a nem egyértelmű kimenetek kék színnel jelennek meg.

Tanulmányunkban az elméleti háttér részletes ismertetésétől eltekintünk (kivéve a homoszkedaszticitás tesztjeit, ahol a felsőoktatásban ritkábban alkalmazott tesztet ismertetjük), mert az az Irodalom részben felsorolt szak-, illetve tankönyvekben, tanulmányokban megtalálható, célunk csupán a szoftver bemutatása, gyakorlati, oktatási célokra való közreadása.

2. Munkalapok

A továbbiakban a munkalapok tartalmát ismertetjük, és mivel a program képes az autokorreláció és a homoszkedaszticitás tesztelésére is, ezért két példán keresztül szemléltetjük a számításokat. Az első példa idősoros adatállomány, a második pedig keresztmetszeti, az adatállományokat elhelyeztük a regresszio.xls parancsfájlban.

2.1. Az Adat munkalap

Az Adat munkalap két nagyobb egységből áll. A bal oldali, sárgával jelölt terület az adatok bevitelére szolgál, itt kell rögzíteni az aktuális adatállományt. Új adatok bevitele előtt a megjelenő mintafeladat adatállományát az Adatok törlése gombra való kattintással törölhetjük. Az új adatokat kell beilleszteni annak érdekében, hogy a

⁶ A magyarázóváltozókra érvényes korlát az Excel sajátja. A megfigyelésekre vonatkozó korlát igény szerint bővíthető, a korlátozás oka a gyors számítási sebesség megtartása.

parancsfájl formátuma megmaradjon. A jobb oldali egység a regressziós modell alapstatisztikáit közli:

- Regressziós statisztika: R – többszörös korrelációs együttható; R^2 – többszörös determinációs együttható; \tilde{R}^2 – korrigált determinációs együttható; s – modell standard hibája; n – megfigyelések száma.
- Varianciaanalízis: a többváltozós regressziós modell varianciaanalízis táblája.
- Regressziós együtthatók: együtthatók értékei és standard hibái, t -értékei, p -értékei, valamint konfidencia intervallumai (tetszőleges megbízhatósági szinten); változók bevonásáról/kihagyásáról döntő jelölőnégyzetek.

A formátum követi az Excel adatelemző menüpontja által használtat, azzal a különbséggel, hogy az egyes cellák függvényeket tartalmaznak, így az adatok megváltozásának hatása azonnal nyomon követhető az eredményeken. Szintén eltérés a beépített funkcióhoz képest, hogy az eredeti adatok meghagyása mellett is kihagyhatunk, illetve újra bevonhatunk változókat a paraméterek soraiban található jelölőnégyzetek segítségével.

A varianciaanalízis tábla segítségével a modell globális próbáját végezhetjük el. A hipotézisrendszerről való döntés – didaktikai okokból – két módon is elvégezhető: a tetszőlegesen beállítható szignifikanciaszinthez tartozó kritikus érték, valamint a p -érték alapján.

A gyors parciális tesztelés lehetőséget biztosít a backward eliminációs módszer alkalmazására. A módszer lényege, hogy az első lépésben olyan regressziós függvényt határozunk meg, amely az összes megfigyelt magyarázóváltozót tartalmazza, majd lépésenként kihagyjuk azokat a változókat, amelyek nem járulnak hozzá szignifikánsan a reziduális négyzetösszeg csökkentéséhez. A változók szelektálásához a p -értékeket használjuk: ha ennek értéke magasabb, mint amit megengedünk (például 0,05), akkor elfogadjuk a nullhipotézist, a regressziós paraméter nem különbözik szignifikánsan nullától. Amennyiben több változó p -értéke is magasabb a kívántnál, úgy a legmagasabb értékkel rendelkező változót hagyjuk ki elsőként. Az eliminációt addig folytatjuk, míg valamennyi bevont paraméter szignifikáns nem lesz.

A változók szelektálását természetesen elvégezhetjük a multikollinearitás vagy a homoszkedaszticitás parciális tesztjei alapján is.

2.2. A Mátrix munkalap

A Mátrix munkalapon a többváltozós regressziószámítással kapcsolatos mátrixok, valamint az ezekhez tartozó statisztikák találhatók meg. A mátrixok maximális

mérete a magyarázóváltozók maximális számával van összhangban. A munkalapon megjelenő mátrixok a következők:

– A „C” oszloptól kezdődően rendre: teljes korrelációs mátrix (valamennyi változóra); bevont korrelációs mátrix (a meghagyott magyarázó változókra, ha valamennyi magyarázóváltozó szerepel a végleges modellben, akkor megegyezik az előző mátrix tartalmával); bevont korrelációs mátrix inverze; determinációs együtthatók a teljes adatmátrixra; determinációs együtthatók a bevont adatmátrixra; bevont változók parciális korrelációit tartalmazó háromszögmátrix; $\mathbf{X}^T \mathbf{X}$ mátrix a teljes adathalmazra; $\mathbf{X}^T \mathbf{X}$ mátrix a bevont változókra; $(\mathbf{X}^T \mathbf{X})^{-1}$ a bevont változókra.

– A „V” oszloptól kezdődően rendre: teljes korrelációs mátrixhoz tartozó t -értékek; bevont korrelációs mátrixhoz tartozó t -értékek; bevont magyarázóváltozók korrelációs mátrixának inverze; bevont változók parciális korrelációihoz tartozó t -érték.

– Az „AP” oszloptól kezdődően rendre: teljes adatmátrixra a sajátértékek és sajátvektorok; bevont adatmátrixra a sajátértékek és sajátvektorok; a sajátértékek megoszlási és kumulált megoszlási viszonyszámái; főkomponenssúly-mátrix; főkomponenssúlyok négyzete.

A felsorolt mátrixok közül több önmagában is fontos információt hordoz a regresszióval kapcsolatban, néhány kiszámítása pedig a további vizsgálatok miatt szükséges. Didaktikai okokból mindegyik mátrix bemutatását szükségesnek tartottuk.

2.3. A Maradék munkalap

A Maradék munkalapon az aktuális modell empirikus maradékaiból képzett oszlopvektorok találhatóak meg, valamint lehetőség van becslés, előrejelzés elvégzésére is. A munkalapon található oszlopvektorok a következők:

- \mathbf{y} – a vizsgált eredményváltozó értékeinek vektora;
- $\hat{\mathbf{y}}$ – az eredményváltozó értékeinek becsült vektora, a bevont magyarázó változókkal történő pontbecslés;
- $\hat{\mathbf{y}}^2$ – az eredményváltozó becsült értékeinek négyzete;
- e – empirikus reziduumok (hibatényezők, hibatagok, maradékok) ($e = \mathbf{y} - \hat{\mathbf{y}}$);
- e_{t-p} – az empirikus maradék p -vel ($p = 1, 2, \dots, 12$) késleltetett értékei (p nagyságát az Autokorreláció munkalapon lehet megadni, jellemzően $p = 1$).

Előrejelzést (idősorok esetén), illetve pontbecslést (keresztmetszeti adatbázisok esetén) a „H” oszloptól kezdődően készíthetünk, a sárga mezőkbe a magyarázóváltozók értékeit kell beírni. Technikai okokból valamennyi (bevont és be nem vont) változóhoz értékeket kell megadni, ezekből csak azokat fogja a program figyelembe venni, amelyek a bevont változókhoz tartoznak. Egyszerre maximum 20 becslés, illetve előrejelzés hajtható végre. A helyesen kitöltött magyarázóváltozó-értékekhez tartozó becslült eredményváltozó-érték a „H” oszlopban olvasható le.

2.4. A Multikollinearitás munkalap

A Multikollinearitás munkalap a magyarázóváltozók összefüggésének problémáját vizsgálja.⁷ Az elvégezhető tesztek közül nem építettük be a programba valamennyit, csupán az oktatásban gyakran alkalmazott, általánosan elterjedt próbákat. A beépített tesztek és módszerek a következők:

- a multikollinearitás globális tesztelése: χ^2 – próba; kondícióindexek és kondíciószám (gyökös formula); Petres-féle RED-mutató (Kovács–Petres–Tóth [2004, 2005]);
- a multikollinearitás lokalizálása: parciális korrelációs együtthatók tesztelése; F -próba; VIF-mutató (variancia infláló faktor); tolerancia-mutató; a multikollinearitás kiküszöbölése: főkomponens regresszió.⁸

2.5. Az Autokorreláció munkalap

Az autokorreláció (Kovács I. [1977] 605. old.) mértékét a reziduális autokorrelációs együtthatóval mérhetjük. A p -ed rendű (p időegységgel késleltetett, a parancsfájl esetében $p = 1, 2, \dots, 12$) elméleti autokorrelációs együtthatót az egymástól p időegységnyi távolságra álló maradéktagok korrelációs együtthatójaként becsülhetjük.

A gyakorlatban az elsőrendű autokorrelációs együtthatót ($p = 1$) szoktuk tesztelni. A fájlban közöltünk több késleltetésre vonatkozó adatot is, amire például szezonaritást mutató adatsorok esetén lehet szükség. Az Excel-parancsfájlban az eredeti, és nem a közelítő p -ed rendű reziduális autokorrelációs együtthatóval számoltunk, majd azt Student-féle t -próba felhasználásával teszteltük. A program kiszámítja a Durbin–Watson-mutató közelítő értékét és teszteli is azt. A kritikus érté-

⁷ A multikollinearitás témaköre jelentős irodalommal rendelkezik, a regressziós modellek becslése és alkalmazása során jelentkező probléma legfrissebb magyar nyelvű összefoglalóját Kovács Péter [2008a] írása adja.

⁸ A számítások elvégzéséhez szükség van a mátrix.xls parancsfájltra.

keket az Excel nem szolgáltatja, így azokat 1 és 5 százalékos szignifikanciaszintekre táblázatból keresi ki a program.

Általánosságban elmondhatjuk, hogy az autokorreláció jelenléte mellett készített paraméter- és pontbecslések ugyan torzítatlanok maradnak, de nem lesznek hatásosak. Különösen óvatosan kell kezelni az autokorrelált modellt, ha segítségével előrejelzéseket kívánunk készíteni. Autokorrelált modellek esetében az együttthatók standard hibái torzítottak, így sem a standard hibákhoz kapcsolódó próbák, sem az előrejelzésekhez kapcsolódó konfidenciaintervallumok nem használhatók fel.

A program ábrázolja e_{t-p} függvényében az e_t alakulását. Az ábra alapján vizuálisan is következtethetünk az autokorreláció léteire, illetve hiányára.

2.6. A Homoszkedaszticitás munkalap

A keresztmetszeti adatok esetében a hibatéyző varianciájának állandóságát tesztjük. Ha konstans a hibatéyző varianciájának várható értéke, akkor:

$$E(\varepsilon_i^2) = \sigma^2 \quad i = 1, 2, \dots, n.$$

Keresztmetszeti adatok esetén homoszkedaszticitás szempontjából is tesztelnünk kell a modelleket, hiszen elméleti feltétel, hogy a hibatéyző varianciája állandó legyen (*Pintér* [1991] 18. old.).

A nullhipotézis:

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2.$$

Az alternatív hipotézis:

$$H_1 : \sigma_l^2 \neq \sigma_m^2,$$

ahol $l, m = 1, 2, \dots, n$ ($l \neq m$).

A nullhipotézis azt fogalmazza meg, hogy a hibatéyző szórásnégyzetei (varianciái) állandók. A nullhipotézis teljesülése egyben azt is jelenti, hogy a modell homoszkedasztikus, az alternatív hipotézis a heteroszkedaszticitás feltételezését szimbolizálja. A heteroszkedaszticitás jelensége esetén a regressziós együttthatók becslése torzítatlan, ugyanis továbbra is feltesszük, hogy a hibatéyző várható értéke nulla. Ugyanakkor a paraméterek varianciájára vonatkozó becslés nem lesz hatásos,⁹ a paraméterek standard hibái torzítottak, használatuk megkérdőjelezhető, a se-

⁹ Ez azt jelenti, hogy a klasszikus legkisebb négyzetek módszere (KLN, Ordinary Least Squares – OLS) alkalmazása esetén a becslések ebben az esetben nem lesznek hatásosak, vagyis található egy másik torzítatlan lineáris becslés, aminek kisebb a varianciája, mint az KLN (OLS)-becslésnek (*Ramanathan* [2003] 365–366. és 397–398. old.)

gítségükkel elvégzett próbák (például t - és F -próbák) és becslések félreinformálhatnának.

2.6.1. Globális (csoportos) BPG, Glejser és KB-próba

A Breusch–Pagan–Godfrey (BPG) és a Glejser-próba esetében a nullhipotézis megegyezik az előzőekben leírtakkal, a hipotézisrendszer általánosabb formában (Glejser [1969] 316–323. old., Godfrey [1978] 227–236. old., Breusch–Pagan [1979] 1287–1294. old., továbbá Ramanathan [2003] 367–369. old., Pintér [1991] 21–24. old., Gujarati [2003] 411–412. old., Maddala [2004] 244–246. old.):

$$H_0 : \sigma_1^2 = \dots = \sigma_n^2$$

$$H_1 : E[f(\varepsilon_i)] = \sigma^2 [h(\mathbf{Z}\boldsymbol{\alpha} + \mathbf{v})],$$

ahol

f – az eredeti reziduumok függvénye (például abszolút értéke, négyzete, logaritmus);

h – a magyarázóváltozók függvénye (a függvény alakja lineáris, hatványkitevős, exponenciális);

\mathbf{Z} – a heteroszkedaszticitást magyarázó változók $n \times (k+1)$ típusú mátrixa;

$\boldsymbol{\alpha}$ – a véletlent becslő modell $(k+1) \times 1$ típusú paramétervektora;

\mathbf{v} – $n \times 1$ típusú, véletlen elemeket tartalmazó vektor.

F -próbával teszteljük a nullhipotézist, aminek elfogadása esetén a modell homoszkedasztikus, elutasítása esetén pedig heteroszkedasztikus.

A globális próbák a következők:

– Glejser-próba:

$$|e_i| = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + v_i.$$

A regresszio.xls fájlban a pótlólagos regresszió többszörös determinációs együtthatója: $R^2(|e_i|; x)$.

– Breusch–Pagan–Godfrey (BPG)-próba:

$$e_i^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + v_i.$$

A regresszio.xls fájlban a pótlólagos regresszió többszörös determinációs együtthatója: $R^2(e^2; x)$.

– Koenker–Bassett (KB)-próba (Gujarati [2003]):

$$e_i^2 = \alpha_0 + \alpha_1 \hat{y}_i^2 + v_i.$$

A regresszio.xls fájlban a pótlólagos regresszió többszörös determinációs együtthatója $R^2(e^2; \hat{y}^2)$.

A képletek jelölései:

k – az eredeti regressziós függvényben a magyarázóváltozók száma;

$i = 1, 2, \dots, n$: a megfigyelések száma;

$|e_i|$ – az eredeti modell reziduális változójának abszolút értéke;

e_i^2 – az eredeti modell reziduális változójának négyzete;

\hat{y}_i^2 – az eredeti függvénnyel becsült eredményváltozó négyzete;

α_j – a becsült paraméterek ($j = 0, 1, 2, \dots, k$);

v_i – a pótlólagos regresszió reziduális változója.

A regresszió paramétereinek együttes szignifikanciája a globális F -próba segítségével mindegyik bemutatott teszt esetében vizsgálható.

2.6.2. A homo- és heteroszkedaszticitás vizsgálata

A Glejser- és a BPG-próba lehetővé teszi a heteroszkedaszticitás lokalizálását. Amennyiben feltételezzük, hogy a magyarázóváltozók lineáris függvényei a reziduális változók abszolút értékei vagy a négyzetei, akkor felírható magyarázóváltozónként egy-egy pótlólagos regressziós egyenlet.

A pótlólagos, j -edik magyarázóváltozóra vonatkozó regressziós egyenletek a következők:

– Glejser-próba esetén:

$$|e_i| = \alpha_0 + \alpha_1 x_{ji} + v_i,$$

– BPG-próba esetén:

$$e_i^2 = \alpha_0 + \alpha_1 x_{ji} + v_i,$$

ahol x_{ji} a j -edik magyarázóváltozó i -edik értéke.

A regressziós együtthatót (meghatározó szerepe az α_1 együtthatónak van) a Student-féle t -próbával teszteljük.

3. Gyakorlati alkalmazások bemutatása idősoros és keresztmetszeti adatok alapján

A regresszio.xls parancsfájl minden esetben közli az Autokorreláció és a Homoszkedaszticitás munkalapokon a számításokat. Az autokorreláció idősoros adatoknál jelentkezik, ahol az adatok sorrendje kötött. A keresztmetszeti adatok sorrendje változtatható, ebben az esetben a homoszkedaszticitást szoktuk vizsgálni. Megjegyezzük, hogy keresztmetszeti adatoknál is előfordul, hogy a szomszédos hibatagok korrelálnak egymással, amit térbeli korrelációnak neveznek. Az autokorreláció vizsgálatánál az ökonometriai szakirodalomban ettől eltekintenek, kizárólag az idősorok hibatagjainak vizsgálata tartozik e témakörbe (*Maddala* [2004] 273–274. old., *Ramanathan* [2003] 361–363. és 399–400. old.; *Gujarati* [2003] 401–403. és 441–443. old.) A maradékváltozó (reziduális változó) vizsgálatánál tehát lényeges kérdés, hogy idősoros vagy keresztmetszeti adatokkal dolgozunk-e. Idősoros adatbázis esetén az autokorrelációt, míg keresztmetszeti adatoknál a homoszkedaszticitást teszteljük. Ennek megfelelően két példát mutatunk be, mindkettő valós magyarországi adatokat tartalmaz.

1. Idősoros példa

Az 1985 és 2008 közötti magyarországi cementtermelést és az azt befolyásoló tényezőket vizsgálatuk.¹⁰ A regressziós modell változói: y – cementtermelés (ezer tonna), x_1 – GDP volumenindexe (1985 = 100 százalék); x_2 – épített lakások száma (darab) – x_3 – építőanyag-ipar volumenindexe (1985 = 100 százalék); x_4 – népesség száma (ezer fő).

A rendszerváltozás idején a hazai cement-előállítás megközelítette az évi négy-millió tonnát, ezt követően azonban drasztikusan visszaesett, és 2000-ig közel egymillió tonnával alatta maradt a csúcsévek termelésének, majd 2001-től emelkedett ugyan a kibocsátás, de 2008-ban is közel félmillió tonnával maradt el az 1990-es szinthez képest.

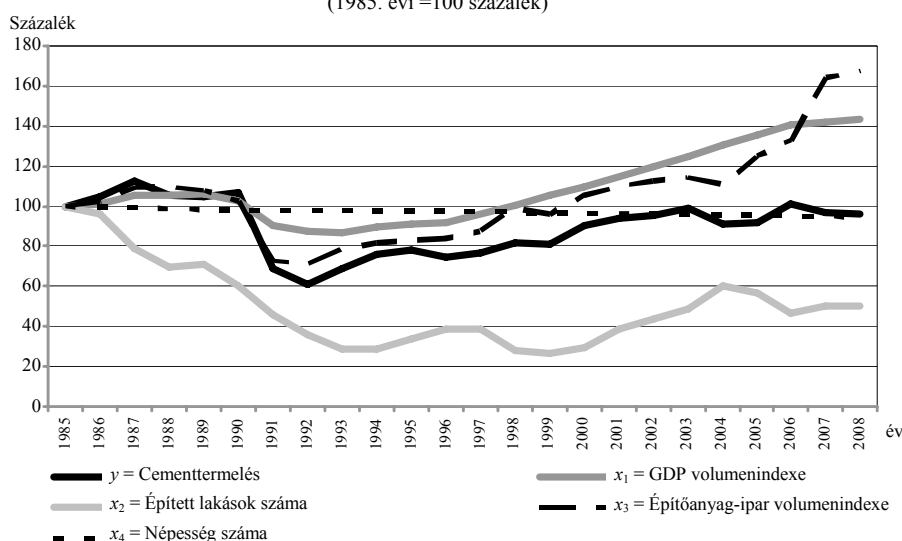
A számítások megkezdése előtt célszerű az adatokat ábrázolni, hogy feltárjuk azok tendenciáit. A cementtermelés és a vizsgált magyarázóváltozók alakulását a következő ábra mutatja. Az ábrakészítés során a vizsgált mutatók arányosságának biztosítása érdekében mindegyik mutatót (tehát a cementtermelést, az épített lakások számát és a népességszámot is) 1985-ös bázison számítottuk.

Az ábra alapján látható, hogy a cementtermelés és a vizsgált magyarázóváltozók sok tekintetben hasonlóan mozognak. A termelés mélypontját a rendszerváltást köve-

¹⁰ Az adatok forrásai: *Polt* [2005] 996. old.; *Hunyadi-Vita* [2008] II. köt. 204. old., CD-melléklet: Adatok8.xls; *KSH* [1985–2005]; *KSH* [1985–2008].

tő években érte el. A magyarázóváltozók közül a népességszám eltérően alakult: Magyarországon a vizsgált időszakban a népességszám folyamatosan csökkent, aminek mértéke a huszonnégy év alatt $-5,2$ százalék volt. Eltérést mutat az épített lakások számának alakulása is, amely 1985 óta csökkenő tendenciát mutat, kivéve az 1995 és 1997, valamint a 2000 és 2003 közötti időszakot.

1. ábra. A cementtermelés és az azt befolyásoló tényezők alakulása Magyarországon 1985 és 2008 között (1985. évi =100 százalék)



Vizsgálhatjuk a ciklusok fordulópontjait is, az átlagos periódushossz¹¹ a cementtermelésnél 3, a GDP volumenindexénél 10, az épített lakások számánál 6, az építőanyag-ipar volumenindexénél 5 év. A népességszám esetében nem voltak fordulópontok.

A termelés elemzése és előrejelzése a regressziószámítás felhasználásával a cementipar esetében arra épült (*Polt* [2005] 996–1000. old.), hogy az építőanyagok és ezen belül a cement termelése, szorosan követi a GDP változását, valamint függhet az épített lakások számának, az építőanyag-ipar teljesítményének és a népesség számának alakulásától is. A népességszám változása és az épített lakások száma közötti kapcsolatot az Egyesült Államok adatbázisán először Kuznets modellezte, kidolgozva a róla elnevezett 15–25 éves építési ciklus elméletét (*Kuznets* [1930]). Az építőanyagok és ezen belül a cement felhasználását az elmúlt években elsődlegesen az építési piac alakulása, pontosabban az infrastruktúra- (autópályák) és a lakásépítés befolyásolta.

¹¹ A ciklusfordulópontok számítása Excel-parancsfájl felhasználásával.

A számítások eredményei

Varianciaanalízis:

Összetevő	df	SS	MS	F-érték	p-érték
Regresszió	4	4845009,2	1211252,3	17,1	0,000004
Maradék	19	1347194,2	70905,0		
Összesen	23	6192203,4			

A varianciaanalízis tábla alapján a nullhipotézist elutasítjuk, tehát van legalább egy olyan magyarázóváltozó, amely szignifikáns hatással rendelkezik, létezik legalább egy nullától eltérő értékű regressziós paraméter.

Regressziós együtthatók:

Együttható	Érték	Standard hiba	t-érték	p-érték	Alsó 95%	Felső 95%
b_0	-54913,67	22223,78	-2,47	0,0231	-101428,57	-8398,77
b_1	49,15	20,68	2,38	0,0281	5,88	92,42
b_2	-0,01	0,01	-0,89	0,3820	-0,04	0,02
b_3	1,23	6,91	0,18	0,8606	-13,24	15,70
b_4	5,16	2,04	2,54	0,0201	0,90	9,43

A regressziós paraméterek parciális tesztelése: a backward eliminációs módszer alkalmazása alapján először mind a négy magyarázóváltozót bevontuk a modellbe, majd az így meghatározott regressziófüggvényből szelektáltuk azokat a változókat, amelyek nem járulnak hozzá szignifikánsan a reziduális négyzetösszeg csökkenéséhez (Mundruczó [1981] 117–118. old.). A változók szelektálásához a p-értékeket használtuk. Ennek alapján először az x_3 változót, majd az x_2 magyarázóváltozót hagytuk ki a modellből. Meg kívánjuk jegyezni, hogy szakmailag indokolt lenne a modellben szerepeltetni a két kihagyott változót.

Regressziós együtthatók:

Együttható	Érték	Standard hiba	t-érték	p-érték	Alsó 95%	Felső 95%
b_0	-37518,27	5884,69	-6,38	0,0000	-49756,15	-25280,39
b_1	38,65	4,62	8,36	0,0000	29,03	48,27
b_4	3,56	0,53	6,67	0,0000	2,45	4,67

A regressziófüggvény tehát:

$$\hat{y} = -37518,27 + 38,65x_1 + 3,56x_4.$$

A multikollinearitás tesztjei:

A χ^2 globális próba alapján 5 százalékos szignifikanciaszinten van multikollinearitás ($\chi^2 = 8,18$; $df = 1$; $p = 0,0042$).

A parciális korrelációs együtthatók alapján számított t -statisztika értéke $-11,66$, a kritikus érték pedig $2,08$, a két magyarázóváltozó között van multikollinearitás.

A p -értékek is a multikollinearitás létét igazolják. A VIF-mutató értéke $2-5$ között van, tehát erős, zavaró a multikollinearitás mértéke.

	y	R^2	F -érték	p -érték	VIF_j	T_j
R^2x	x_1	0,584	30,86	0,0000	2,40	0,42
R^2x	x_4	0,584	30,86	0,0000	2,40	0,42

A kondícióindexek (CI – condition index) a magyarázóváltozók korrelációs mátrixának legnagyobb (λ_{\max}) és j -edik (λ_j) $j = 1, 2, \dots, k$ sajátértékei alapján határozhatók meg (Kotz et al. [2006] 1239–1240. old.):

$$CI = \sqrt{\frac{\lambda_{\max}}{\lambda_j}}$$

Ha a legkisebb sajátértéket λ_{\min} -nel jelöljük, akkor a kondíciószám (CN – condition number):

$$CN = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

Ha a magyarázóváltozók lineárisan függetlenek, valamennyi sajátérték egy, akkor a CN -mutató értéke is eggyel egyenlő. Minél nagyobb a mutató, annál erősebb a multikollinearitás mértéke. A multikollinearitás mértéke gyenge, ha $1 < CN < 5$, zavaró, ha $5 < CN < 10$, igen zavaró, ha $CN > 10$.

Esetünkben a kondíciószám $2,734$, azaz a mutató szerint gyenge multikollinearitást tapasztalunk a két magyarázóváltozó között.

A Petres-féle RED-mutatót is számszerűsítettük, $76,4$ százalékos eredményt kaptunk, a kritikus érték pedig 100 százalék. Ha minden sajátérték egy, akkor $RED(\%) = 0\%$. Ez azt jelenti, hogy a sajátértékek szorzata, vagyis a magyarázóváltozók korrelációs mátrixának a determinánsa eggyel egyenlő. Ebben az esetben a mátrix ortogonális, nincs multikollinearitás, a magyarázóváltozók függetlenek egymástól. Amennyiben a sajátértékek távolodnak ettől az esettől, akkor a RED-mutató értéke növekszik. A maximális redundancia esetén a mutató értéke 100 százalék.

Ha a számított érték a kritikusknál kisebb, akkor a lineáris regressziós modell illesztése után kapott becült paraméterek szórásnégyzeteinek az összege, illetve átlaga biztosan véges. Ellenkező esetben a lineáris regressziós modell illesztése után kapott becült paraméterek szórásnégyzeteinek az összege, illetve átlaga nem biztos, hogy véges, az adatállomány redundáns.

Esetünkben az adatállomány a Petres-féle RED-mutató alapján nem redundáns.

Az autokorreláció tesztelése:

Az elsőrendű reziduális autokorrelációs együttható alapján nincs szignifikáns autokorreláció a modellben:

Autokorreláció rendje	ρ	t	t_{krit}	p -érték
1	0,342	1,707	2,074	0,1018

A népesség a vizsgált időszakban végig csökkent, a GDP volumenindexe pedig – a rendszerváltást követő éveket leszámítva – növekvő trendet mutatott, ezért a két magyarázóváltozó együttes alkalmazása multikollinearitást okozott. Az optimális regressziós egyenes meghatározásához ezért más megoldást kellett keresnünk.

Szakmai indokok alapján új modellt építettünk, és azt kaptuk, hogy a modell globálisan és parciálisan is elfogadható, ha az x_2 és x_3 változókat vonjuk be a modellbe. Nyilvánvaló, hogy az épített lakások számának és az építőanyag-ipar volumenindexének változása (növekedése vagy csökkenése) a cementfelhasználást, és így a termelést is jelentősen befolyásolja. Természetesen befolyásoló tényező a cement export- és importvolumene, de ennek vizsgálatától eltekintettünk. Megállapítható továbbá, hogy a multikollinearitás mértéke és az autokorreláció nem zavaró.

A varianciaanalízis F -próbájához tartozó p -érték ebben az esetben 0,000002, tehát a nullhipotézist elutasíthatjuk.

Regressziós együtthatók:

Együttható	Érték	Standard hiba	t -érték	p -érték	Alsó 95%	Felső 95%
b_0	1476,00	285,95	5,16	0,0000	881,33	2070,67
b_2	0,0204	0,00	4,86	0,0001	0,01	0,03
b_3	10,2928	2,58	4,00	0,0007	4,94	15,65

$$\hat{y} = 1476 + 0,0204x_2 + 10,2928x_3.$$

A multikollinearitás próbái:

A χ^2 globális próba alapján 5 százalékos szignifikanciaszinten nincs multikollinearitás ($\chi^2 = 0,49$; $df = 1$; $p = 0,4821$).

A parciális korrelációs együtthatók alapján számított t -statisztika $-1,77$. A kritikus érték 5 százalékos szignifikanciaszinten 2,08, tehát a két magyarázóváltozó között nincs multikollinearitás. A p -értékek is a multikollinearitás hiányát igazolják. A VIF-mutató értéke 1 és 2 között van, tehát nem zavaró a hatás.

y	R^2	F -érték	p -érték	VIF_j	T_j
x_2	0,052	1,20	0,2860	1,05	0,95
x_3	0,052	1,20	0,2860	1,05	0,95

A kondíciós szám esetünkben 1,26, ami gyenge multikollinearitásra utal.

A Petres-féle RED-mutató:

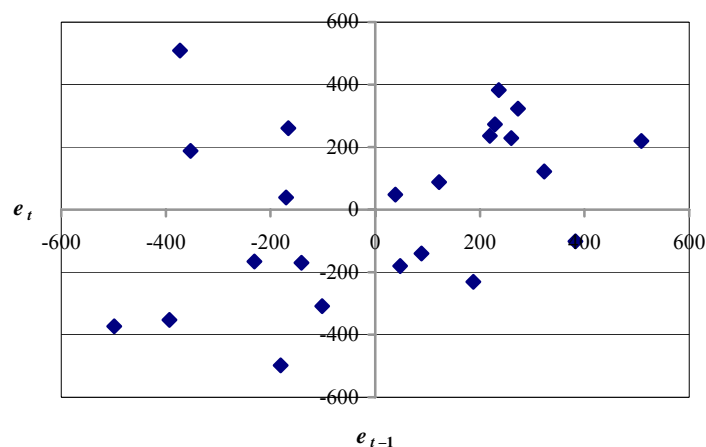
A modell nem redundáns. $RED(\%) = 22,7\%$, ami azt jelenti, hogy az adott méretű és minimális redundanciájú adatállományhoz képest a hasznos tartalmat hordozó adatok aránya 77,3 százalék, azaz az adatok átlagos együttmozgásának a maximumhoz viszonyított mértéke 22,7 százalék.

Az autokorreláció tesztelése:

Az elsőrendű reziduális autokorrelációs együttható alapján nincs szignifikáns autokorreláció a modellben:

Autokorreláció rendje	ρ	t	t_{krit}	p -érték
1	0,365	1,837	2,074	0,0797

2. ábra. Reziduumok ábrája



A Durbin–Watson-féle teszt eredménye: 1,27, ami a bizonytalansági tartományba esik mindkét kérhető szignifikanciaszinten.

A kiválasztott modell az elméleti feltételeknek megfelel, elemzésre és előrejelzésre felhasználható.

2. Keresztmetszeti adatokon alapuló példa

A keresztmetszeti adatok alapján történő regressziószámítást egy tapasztalatiárindex-modellen keresztül mutatjuk be.

Az ökonometriai modellek egyik speciális fajtája a tapasztalati (hedonikus) árindex-modell (*Ramanathan* [2003] 23. old.), amelyben egy árucikk ára a jellemzőitől függ, példa erre a gépkocsi ára és tulajdonságai közötti összefüggés. A vizsgálatba a 10 millió forintnál olcsóbb, hazai forgalmazású autókat vontuk be. A gépkocsik árát nemcsak mérhető tulajdonságai befolyásolják, hanem minőségi tényezők is, mint például a márka, a biztonság, garancia stb.

A mintafeladatban 119 autó adatait vizsgáltuk 2008. évi áron (forrás: <http://www.auto2.hu/>). A modell változói: y – a termék, az új autók alapárjai (ezer forint); x_j – a termék, az új autók tulajdonságai, az autók árát befolyásoló tényezők.

A magyarázóváltozók a következők: x_1 – KÖBCM hengerűrtartalom (cm³); x_2 – TELJ teljesítmény (LE); x_3 – NYOM maximális nyomaték (Nm); x_4 – GYORS 0-ról 100 km/h-ra gyorsulás ideje (sec); x_5 – VMAX végsebesség (km/h); x_6 – TÖMEG satját tömeg (kg); x_7 MTÖMEG megengedett össztömeg (kg); x_8 – HOSZZ hosszúság (mm); x_9 – SZÉLES szélesség (mm); x_{10} MAGAS magasság (mm); x_{11} FOGYV fogyasztás városban (liter/100 km); x_{12} FOGYVK fogyasztás városon kívül (liter/100 km).

Az autóárak és az autóárakat befolyásoló 12 magyarázóváltozó közötti regressziós kapcsolat vizsgálata alapján a következő fontosabb megállapításokat tehetjük:

- A modell minden számított teszt alapján homoszkedasztikus.

- A modellben minden számított teszt alapján káros mértékű a multikollinearitás. Ennek oka, hogy az autók tulajdonságai közül a teljesítmény erőteljesen befolyásolja a többi magyarázóváltozót (a sebességet, a fogyasztást, a gyorsulást, a végsebességet, a tömeget stb).

- A multikollinearitás miatt a regressziós paraméterek standard hibái nagyobbak (a VIF-mutató például 10 magyarázóváltozó esetében a kritikus értéknél nagyobb), és csak a b_0 és b_3 regressziós paraméter különbözik a t -próba alapján 5 százalékos szignifikanciaszinten nullától.

- Figyelembe véve, hogy mind a 12 magyarázóváltozónak a modellben való megtartása indokolt, célszerű a főkomponens-elemzést (PCA – Principal Components Analysis) elvégezni.

A regresszio.xls program közli a bevont változókra vonatkozó, a számításokhoz szükséges sajátértékeket és sajátvektorokat, továbbá a sajátértékek megoszlási és kumulált megoszlási viszonyszámait.

A főkomponensanalízis-számítások részletei megtalálhatók a <http://www.gmi.ktk.pte.hu/index.php?mid=33#SiposB> oldalon letölthető kézikönyv 124–126. oldalán. A transzformált paramétereket, a számítások végeredményét az alábbi táblázatban mutatjuk be.

Változók	Transzformált paraméterek
x_1	2,099
x_2	2,615
x_3	0,597
x_4	-3,513
x_5	3,204
x_6	0,290
x_7	-0,196
x_8	1,151
x_9	0,382
x_{10}	-3,290
x_{11}	2,894
x_{12}	2,218

4. Összefoglalás

A regresszio.xls program felhasználása nagymértékben segíti a regressziós modellezést, valamint annak oktatását. A különböző magyarázóváltozók kombinálásával kialakítható modellek gyors értékelésére ad módot, az adatállomány tetszőleges változtatására az eredmények minden esetben reagálnak. A magyarázóváltozók számának növekedésével a lehetséges modellvariánsok száma többszöröződik.

Nemcsak a modell globális és parciális tesztelésének az eredményét látjuk azonnal, hanem idősorok esetén az autokorreláció, keresztmetszeti adatoknál pedig a homoszkedaszticitás tesztjeit is értékelhetjük, valamint a reziduum ábrákat elemezhetjük. A magyarázóváltozók összefüggésének vizsgálatára több teszt is lehetőséget ad.

Tanulmányunkban két példán keresztül mutattuk be a kifejlesztett alkalmazást: az idősoros példa alkalmas volt a backward regresszió, valamint a szakmai ismeretek alapján történő modell felállítására is. Keresztmetszeti adatokon a főkomponens-regresszió alkalmazhatóságát mutattuk be. A modellezés során a modell feltételeinek különböző tesztjeit is minden esetben figyelembe kell vennünk.

Irodalom

- ACZEL, A. D. [2002]: *Complete Business Statistics*. McGraw-Hill/Irwin. Boston.
- BACZONI P. [2007]: *Egyszerűen Microsoft Office Excel 2003*. Panem Kiadó. Budapest.
- BALÁZSNÉ MÓCSAI A. – CSETÉNYI A. [2003]: *Kvantitatív technikák, II.* Zsigmond Király Főiskola. Budapest.
- BALOGH M. [2000]: *Statisztikai ismeretek*. Perfekt Kiadó. Budapest.
- BÁRTFAI B. [2002]: *Office XP. World 2002. Exel 2002. Power Point 2002. Outlook Access 2002.* BBS-Info Kft. Budapest.
- BEDŐ, ZS. – RAPPAL, G. [2006]: Is There Causal Relationship Between the Value of the News and Stock Returns? *Hungarian Statistical Review*. Special Number 10. 81–99. old.
http://www.ksh.hu/statszemle_archive/2006/2006_K10/2006_K10_081.pdf (Elérés dátuma: 2010. május 18.)
- BELSLEY, D. A. – KUH, E. – WELSCH, R. E. [1982]: *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley and Sons. New York.
- BERENSON, M. L. – LEVINE, D. M. – KREHBIEL, T. C. [2006]: *Basic Business Statistics: Concepts and Applications*. Pearson/Prentice Hall. New Jersey.
- BESENYEI L. – GIDAI E. – NOVÁKY E. [1977]: *Jövő kutatás, előrejelzés a gyakorlatban*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- BREUSCH, T. S. – PAGAN, A. R. [1979]: Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* (Econometric Society). 47. évf. 5. sz. 1287–1294. old.
- EVANS, J. R. [2007]: *Statistics, Data Analysis, and Decision Modeling*. Pearson-Prentice Hall. New Jersey.
- FARRAR, D. E. – GLAUBER, R. R. [1967]: Multicollinearity in Regression Analysis: The Problem Revisited. *Review of Economics and Statistics*. 49. évf. 1. sz. 92–107. old.
- GLEJSER, H. [1969]: A New Test for Heteroscedasticity. *Journal of the American Statistical Association*. 64. évf. 325. sz. 316–323. old.
- GODFREY, L. [1978]: Testing for Multiplicative Heteroscedasticity. *Journal of the American Statistical Association*. 8. évf. 2. sz. 227–236. old.
- GOLDFELD, S. M. – QUANDT, R. E. [1965]: Some Tests for Homoscedasticity. *Journal of the American Statistical Association*. 60. évf. 310. sz. 539–547. old.
- GREENE, W. H. [2003]: *Econometric Analysis*. Pearson Education International. Upper Saddle River. Prentice Hall. New Jersey.
- GUJARATI, D. N. [2003]: *Basic Econometrics*. McGraw-Hill Higher Education.
- HAJDU O. – HUNYADI L. [1995]: Varianciafelbontás: előfeltevések és következtetések. *Sigma*. 1–2. sz. 1–18. old.
- HAJDU O. ET AL. [1987]: *Ökonometriai alapvetés*. Tankönyvkiadó. Budapest.
- HAJDU O. ET AL. [1994–1995]: *Statisztika I-II*. JPTE Kiadó. Pécs.
- HAJDU O. [2003]: *Többváltozós statisztikai számítások*. Központi Statisztikai Hivatal. Budapest.
- HARRISON, M. J. [1982]: Tables of Critical Values for a Beta Approximation to Szroeter's Statistic for Testing for Heteroscedasticity. *Oxford Bulletin of Economics and Statistics*. 44. évf. 2. sz. 159–167. old.
- HARVEY, A. C. [1976]: Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica*. 44. évf. 3. sz. 461–466. old.

- HARVEY, G. [2000]: *Excel 2000 for Windows for Dummies*. Kossuth Kiadó. Budapest.
- HILL, R. C. – GRIFFITHS, W. E. – LIM, G. C. [2008]: *Principles of Econometrics*. John Wiley and Sons. New York.
- HUNYADI L. – VITA L. [2002]: *Statisztika közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- HUNYADI L. – VITA L. [2008]: *Statisztika I-II*. Aula Kiadó. Budapest.
- HUNYADI L. [2001]: *Statisztikai következtésemélet közgazdászoknak*. Központi Statisztikai Hivatal. Budapest.
- HUNYADI L. [2006]: A heteroszkedaszticitásról egyszerűbben. *Statisztikai Szemle*. 84. évf. 1. sz. 75–82. old. http://www.ksh.hu/statszemle_archive/2006/2006_01/2006_01_075.pdf (Elérés dátuma: 2010. május 18.)
- JÁNOSA A. [2005]: *Adatelemzés számítógéppel*. Perfekt Kiadó. Budapest.
- KÁDAS K. [1944]: Az emberi munka termelékenységének statisztikai vizsgálata a magyar gyáriparban. (A Cobb-Douglas féle statisztikai törvény kiegészítése.) *Magyar Statisztikai Szemle*. 22. évf. 7–8. sz. 270–318. old. http://www.ksh.hu/statszemle_archive/viewer.html?ev=1944&szam=07-08&old=3&lap=46 (Elérés dátuma: 2010. május 18.)
- KEHL D. – SIPOS B. [2009]: A telítődési, a logisztikus és életgörbe alakú trendfüggvények becslése Excel parancsfájl segítségével. *Statisztikai Szemle*. 87. évf. 4. sz. 381–411. old. http://www.ksh.hu/statszemle_archive/2009/2009_04/2009_04_381.pdf (Elérés dátuma: 2010. május 18.)
- KERÉKGYÁRTÓ GY. – MUNDRUCZÓ GY. – SUGÁR A. [2001]: *Statisztikai módszerek és alkalmazásuk a gazdasági, üzleti elemzésekben*. Aula Kiadó. Budapest.
- KERÉKGYÁRTÓ GY. – MUNDRUCZÓ GY. [1995]: *Statisztikai módszerek a gazdasági elemzésben*. Aula Kiadó. Budapest.
- KING M. L. [1981]: A Note on Szroeter's Bound Test. *Oxford Bulletin of Economics and Statistics*. 43. évf. 3. sz. 315–322. old.
- KNÜSEL L. [1998]: On the Accuracy of Statistical Distributions in Microsoft Excel 97. *Computational Statistics and Data Analysis*. 26. évf. 3. sz. 375–377. old.
- KNÜSEL L. [2002]: On the Reliability of Microsoft Excel XP for Statistical Purposes. *Computational Statistics and Data Analysis*. 39. évf. 1. sz. 109–115. old.
- KNÜSEL L. [2005]: On the Accuracy of Statistical Distributions in Microsoft Excel 2003. *Computational Statistics and Data Analysis*. 48. évf. 3. sz. 445–449. old.
- KOOP, G. [2008]: *Közgazdasági adatok elemzése*. Osiris Kiadó. Budapest.
- KÖRÖSI G. – MÁTYÁS L. – SZÉKELY I. [1990]: *Gyakorlati ökonometria*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- KOTZ, S. ET AL. [2006]: *Encyclopedia of Statistical Sciences*. 16 Volume Set. John Wiley and Sons. New York.
- KOVÁCS I. [1977]: Az autokorreláció vizsgálata regressziós modellekben. *Statisztikai Szemle*. 552. évf. 6. sz. 603–621. old. http://www.ksh.hu/statszemle_archive/viewer.html?ev=1977&szam=06&old=45&lap=19 (Elérés dátuma: 2010. május 18.)
- KOVÁCS P. – PETRES T. – TÓTH L. [2004]: Adatállományok redundanciájának mérése. *Statisztikai Szemle*. 82. évf. 6–7. sz. 595–604. old.

- http://www.ksh.hu/statszemle_archive/2004/2004_06-07/2004_06-07_595.pdf (Elérés dátuma: 2010. május 18.)
- KOVÁCS P. [2008a]: A multikollinearitás vizsgálata lineáris regressziós modellekben. *Statisztikai szemle*. 86. évf. 1. sz. 38–67. old.
http://www.ksh.hu/statszemle_archive/2008/2008_01/2008_01_038.pdf (Elérés dátuma: 2010. május 18.)
- KOVÁCS P. [2008b]: A statisztikaoktatás módszertanának modernizálása? *Statisztikai Szemle*. 86. évf. 12. sz. 1143–1157. old.
http://www.ksh.hu/statszemle_archive/2008/2008_12/2008_12_1143.pdf (Elérés dátuma: 2010. május 18.)
- KOVACS, P. – PETRES, T. – TOTH, L. [2005]: A New Measure of Multicollinearity in Linear Regression Models. *International Statistical Review*. 73. évf. 3. sz. 405–412. old.
- KOVALCSIK G. [2000]: *Excel 2000*. ComputerBooks. Budapest.
- KREKÓ B. [1966]: *Mátrixszámítás*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- KSH (KÖZPONTI STATISZTIKAI HIVATAL) [1985–2005]: *Ipari és építőipari statisztikai évkönyv*. Budapest.
- KSH (KÖZPONTI STATISZTIKAI HIVATAL) [1985–2008]: *Magyar statisztikai évkönyv*. Budapest.
- KUZNETS, S. S. [1930]: *Secular Movements in Production and Prices: Their Nature and Bearing upon Cyclical Fluctuations*. Houghton-Mifflin. Boston.
- LÉNÁRT I. – RAPPAL G. [2001]: Néhány gondolat a varianciabecslés hibahatáráról. *Statisztikai Szemle*. 79. évf. 7. sz. 613–621. old.
http://www.ksh.hu/statszemle_archive/2001/2001_07/2001_07_613.pdf (Elérés dátuma: 2010. május 18.)
- MADDALA, G. S. [2004]: *Bevezetés az ökonometriába*. Nemzeti Tankönyvkiadó. Budapest.
- MCCULLOUGH B. D. – WILSON B. [2002]: On the Accuracy of Statistical Procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics and Data Analysis*. 40. évf. 4. sz. 713–721. old.
- MCCULLOUGH, B. D. – WILSON, B. [1999]: On the Accuracy of Statistical Procedures in Microsoft EXCEL 97. *Computational Statistics and Data Analysis*. 31. évf. 1. sz. 27–37. old.
- MUNDRUCZÓ GY. [1981]: *Alkalmazott regressziószámítás*. Akadémiai Kiadó. Budapest.
- MUNDRUCZÓ GY. [1998]: *Útmutató a statisztikai modellezéshez*. Aula Kiadó. Budapest.
- NASH, J. C. [2008]: Teaching Statistics with Excel 2007 and Other Spreadsheets. *Computational Statistics and Data Analysis*. 52. évf. 10. sz. 4602–4606. old.
- NYITRAI F.-NÉ – RÉDEY K. [1974]: *Statisztika III*. (Korszerű statisztikai módszerek és alkalmazásuk a gyakorlati közgazdasági munkában). Tankönyvkiadó. Budapest.
- PARK, R. E. [1966]: Estimation with Heteroscedastic Error Terms. *Econometrica*. 34. évf. 4. sz. 888. old.
- PAWLOWSKI Z. [1970]: *Ökonometria*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- PÉTERY K. [2003]: *Táblázatkezelés Excel 2002*. Kossuth Kiadó. Budapest.
- PINTER J. – RAPPAL G. (szerk.) [2007]: *Statisztika*. Pécsi Tudományegyetem, Közgazdaságtudományi Kar. Pécs.
- PINTER J. [1991]: A heteroszkedaszticitás diagnosztizálása. *Statisztikai Szemle*. 69. évf. 1. sz. 16–36. old.

- http://www.ksh.hu/statszemle_archive/viewer.html?ev=1991&szam=01&old=18&lap=21
(Elérés dátuma: 2010. május 18.)
- PINTÉR J. [2000]: *Bevezetés a statisztika módszereibe*. Pécsi Tudományegyetem. Pécs.
- POLT R. [2005]: Levegőkereskedelem – a Nemzeti Kiosztási Terv kialakítása. *Statisztikai Szemle*. 83. évf. 10–11. sz. 990–1009. old.
http://www.ksh.hu/statszemle_archive/2005/2005_10-11/2005_10-11_990.pdf (Elérés dátuma: 2010. május 18.)
- RAMANATHAN, R. [2003]: *Bevezetés az ökonometriába alkalmazásokkal*. Panem Kiadó. Budapest.
- RAPPAI G. [2001]: *Üzleti statisztika Excellel*. Központi Statisztikai Hivatal. Budapest.
- RAPPAI G. [2008]: Gondolatok a gazdaságtudományi képzési területen folyó statisztikaoktatásról. *Statisztikai Szemle*. 86. évf. 9. sz. 829–849. old.
- SAVIN, N. E. – WHITE, K. J. [1977]: The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica*. 45. évf. 8. sz. 1989–1996. old.
- SIPOS B. [1982]: *Termelési függvények – vállalati prognózisok*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- SIPOS, B. – KISS, T. [1998]: REGAL: Expert System for Multiple Linear Regression Analysis. *Hungarian Statistical Review*. Special Number 2. 35–49. old.
http://www.ksh.hu/statszemle_archive/1998/1998_K2/1998_K2_035.pdf (Elérés dátuma: 2010. május 18.)
- SPIEGEL M. R. [1995]: *Statisztika – elmélet és gyakorlat (SI mértékegységekkel)*. Panem–McGraw-Hill. Budapest.
- STUART, A. – ORD, J. K. [1994]: *Kendall's Advanced Theory of Statistics*. Vol. 1. Distribution Theory. Edward Arnold. London.
- SYDSAETER, K. – HAMMOND, P. I. [2006]: *Matematika közgazdászoknak*. Aula Kiadó. Budapest.
- SZROETER, J. [1978]: A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models. *Econometrica*. 46. évf. 6. sz. 1311–1327. old.
- THEISS E. (szerk.) [1958]: *Korreláció és trendszámítás*. Közgazdasági és Jogi Könyvkiadó. Budapest.
- YALTA, A. T. [2008]: The Accuracy of Statistical Distributions in Microsoft® Excel 2007. *Computational Statistics and Data Analysis*. 52. évf. 10. sz. 4579–4586. old.

Summary

The Excel environment of the regression file developed by the authors is accessible for almost anyone, which helps the proliferation of further applications and its usability in higher education. Models built by the manipulation of explanatory variables, as well as the dataset itself can be evaluated rapidly. The authors have developed numerous Excel files also in the fields of special regression analysis, primarily in time series decomposition model applications. All files and the manual are accessible online.