

Statisztikai adatok, anomáliák, csalásfelderítés

Néhány gondolat Tátrai Annamária és Gábos András "Kutatói aggodalmak az EU-SILC magyarországi jövedelemadataival kapcsolatban" című tanulmányához

Tóth István János

Donations from Hungarian citizens and companies and volunteer work by Hungarian and foreign citizens made the research possible.

The IT support partner:

3gteam ltd.: <http://www.3gteam.hu/>

Staff:

Miklós Hajdu	research fellow
Balázs Molnár	IT specialist
Nóra Kovács	intern
István János Tóth	director
Márton Vida	intern

Voluntary Experts:

Katalin Andor	economist
Katalin Goldstein	English language consultant
Jenő Gyenese	software engineer
Dr. Magda József	lawyer, public procurement specialist
Dr. Zoltán Kelemen	lawyer, public procurement specialist
Tibor Kiss	public relations consultant
Attila Székely	public procurement specialist
Barb West	proofreader and editor

CRCB Nonprofit Ltd.

e-mail: info@crcb.eu

website: <http://www.crcb.eu/>

Recommended citation:

Tóth, I. J. (2025). *Statisztikai adatok, anomáliák, csalásfelderítés. Néhány gondolat Tátrai Annamária és Gábor András "Kutatói aggodalmak az EU-SILC magyarországi jövedelemadataival kapcsolatban" című tanulmányához.* CRCB Research Notes, 2025:1, CRCB. <https://www.crcb.eu/?p=3669>

Date of publication: June 11, 2025

"Minthogy ugyanis ama törvények segítségével az anyag egymásután felvesz minden formát, amire csak képes, e formákat sorba véve, végül eljuthatunk majd ahhoz, amely ennek a világnak a formája; úgyhogy hamis hipotézisekből folyó tévedéstől nem kell félni."

[Spinoza: Descartes Alapelvei, III. rész, 216. oldal]

Tartalom

Summary	5
Bevezetés	7
1. Tátrai és Gábor tanulmányának két eredménye és a KSH.....	9
2. Csalásfelderítés – néhány példa	16
3. Közbeszerzések – egy feltételezett csalás	20
4. Összefoglalás és következtetések	36
Hivatkozások	39
Melléklet.....	45
M1. A Newcomb-Benford-törvénnyel foglalkozó közlemények száma 1881-2024	45
M2. A Newcomb-Benford-törvénnyel foglalkozó közlemények listája 2025. május 20-ig a https://www.benfordonline.net oldalon:.....	46
M3. A Világbank Doing Business rangsorának esete	47

Summary

The analysis uses a hypothetical example to demonstrate the consequences of data manipulation on the descriptive statistics and distribution of a statistical variable. It also illustrates how a method commonly applied in fraud detection—the analysis of digit distributions—can assist in uncovering manipulation.

The relevance of this analysis is underscored by recent findings from two Hungarian researchers, who identified unusual anomalies in the income survey data published by the Hungarian Central Statistical Office (HCSO) within the framework of the EU-SILC survey between 2018 and 2023 (Tátrai & Gábos, 2005a; Élő, 2025a; Tátrai & Gábos, 2005b). They found that in the HCSO income surveys from 2018 to 2023, there were unusually high frequencies of values around the poverty threshold, particularly just above it. This unnatural pattern has not been observed in other EU countries or Hungarian data before 2018.

In the first part of the analysis, we briefly summarize the main findings of the studies by Tátrai and Gábos and present the HCSO's press release in response (KSH, 2025). The second part discusses the application of Benford's Law in fraud detection and its limitations. In the third part, we use a hypothetical example to show how even a small-scale manipulation (affecting only a minor portion of the dataset) can significantly distort the distribution of the variable in question and its digit frequencies while remaining invisible to traditional descriptive statistics. In this example, we apply the last-digit test.

The hypothetical case involves a fictional researcher, *X*, who manipulates a real dataset—the net value of public procurement contracts in Hungary. *X* alters the data by replacing actual contract values below a certain threshold with values above that threshold to reduce the perceived dominance of the Orbán-era kleptocratic system. *X* attempts to conceal the manipulation as much as possible by tampering with only 3.2% of all contracts and using a sophisticated technique informed by Benford's Law.

Because we have access to both the original (untainted) and the manipulated data, it is easy to visualize how the hypothetical manipulation alters the distribution of procurement contracts. The last-digit test reveals in numerical terms that—even though the original dataset also deviated somewhat from Benford's expected distribution—the small-scale manipulation affecting just 3% of the data introduces significant new irregularities.

Furthermore, the example shows that while (a) the fraud affected only 3% of contracts, it (b) still partially fulfilled the fraudster's goal: it increased the proportion of contracts above the threshold won by "other Hungarian firms," thereby making them appear closer to the firms tied to the top actors of the kleptocratic regime.

We summarize the study's findings in the fourth part and draw several conclusions. In the appendix, we briefly present the World Bank's Doing Business ranking case,

whose lessons are worth considering when statistical data from an official source is subject to well-founded criticism.

Based on the findings, we draw four conclusions:

1. **Fraud detection methods based on Benford's Law can help identify anomalies and uncover** their causes. However, the power of these methods should not be overestimated. A skilled fraudster familiar with such techniques may be able to conceal manipulation. However, doing so requires knowledge of Benford's Law and the relevant academic findings. The more effectively a fraudster wants to hide their manipulation, the more they need to understand these methods. This effect, in turn, increases the cost of committing fraud: the more sophisticated the fraud needs to be to remain undetected, the more effort it requires. Consequently, Benford's Law can indirectly help reduce the frequency of fraud by raising its cost.

Furthermore, it is important to note that a deviation from the distribution predicted by Benford's Law does not automatically imply fraud. There may be other legitimate explanations for such a deviation besides manipulation.

2. **Even small-scale fraud (affecting 2–3% of a dataset) can significantly impact** the distribution of a variable and can be detected using Benford-based methods.
3. **The administrative setting of thresholds can have unintended effects.** Actors tend to adapt their behavior to these thresholds to maximize their advantage, leading to clustering values just below or above the set limits. In our hypothetical case, fraudster X aimed to reshape the original distribution to align with a personally defined threshold.
4. **It is a fundamental principle that when microdata is available, researchers should examine the distribution of key variables by checking it with histograms** as an early step in the analysis, in addition to calculating descriptive statistics. In light of the anomalies uncovered by Tátrai and Gábos (2025a; 2025b) and the HCSO's response (KSH, 2025), this requirement is equally relevant for all datasets and indicator provided by the HCSO.

Bevezetés

0.1. Az elemzés¹ egy képzeletbeli példán keresztül mutatja be, hogy hogyan segíthet a csalásfelderítésben használt Newcomb-Benford-törvény² annak eldöntésében, hogy egy adatállomány tartalmaz-e statisztikai anomáliát.

0.2. Az elemzés aktualitását az adja, hogy két magyar kutató szokatlan jelenséget fedezett fel a Központi Statisztikai Hivatal (KSH) által az EU-SILC kutatás keretében publikált lakossági jövedelemfelvételek 2018-2023 közötti adataiban (Tátrai & Gábos, 2005a; Élő, 2025a; Tátrai & Gábos, 2005b)³.

¹ Ezúton szeretném megköszönni a Gábos Andrásnak, Molnár Györgynek, Tátrai Annamáriának és Varga Júliának a tanulmány első változatához fűzött megjegyzéseit és javaslatait.

² Egy adathalmazban az első számjegyre vonatkozóan a Newcomb-Benford, illetve Benford-törvény (Benford, 1938) kimondja, hogy a d (ahol tízes számrendszerben $d \in \{1, 2, \dots, 9\}$) első számjegyként való előfordulásának valószínűsége $P(d) = \log_{10}(1 + 1/d)$. Általánosabban, az első n szignifikáns számjegyekre (például 2.383.436, vagy 567.934.567 számoknál legyen az n az első három számjegy, sorrendben 238 és 567) vonatkozóan és kiterjesztve $B \geq 2$ számrendszerekre: $P(n) = \log_B(1 + 1/n)$ (Miller, 2015). Maga a Benford törvény egyébként a tudománytörténetben nem ritka újrafelfedezés egyik példája, mivel ezt az összefüggést eredetileg Newcomb fedezte fel 57 évvel Benford előtt (Newcomb, 1881). A felfedezés története mindkét esetben figyelemre méltó: Newcomb csillagászként, Benford fizikusként számításaihoz rendszeresen használtak logaritmus táblázatot. Azt figyelték meg, hogy a hosszú használat során a táblázat első lapjai sokkal gyorsabban használódtak el, sérültek meg, mint a táblázat utolsó lapjai. Newcomb nem tulajdonított túl nagy jelentőséget felfedezésének. Megfigyelését egy két oldalas kutatási jegyzetben közölte. Cikkje feledésbe merült, míg Benford, éppen ellenkezőleg - nem tudva Newcomb cikkéről – már fontos felfedezésként kezelte (Benford, 1938). A szakirodalom általában Benford-törvényként hivatkozik rá, de az újabb munkák már megemlítik Newcombot, mint a törvény első felfedezőjét (például Kossovsky, 2015; Miller, 2015) és Newcomb-Benford törvényként is írnak róla. A Newcomb-Benford törvénnyel kapcsolatos publikációk száma 1970-ig elenyésző volt és 2000-ig nem haladta meg a tíz publikációt évente. Az igazi áttörés – feltehetően az elemezhető és elérhető adatsorok számának jelentős növekedése és a statisztikai szoftverek használatának elterjedése nyomán – ezután következett be. Feltehetően Hal Varian cikke is szerepet játszott ebben (Varian, 1972). Varian felvetette, hogy a Benford törvényt lehetne statisztikai adatok ellenőrzésére is használni. Ezzel egyébként egy speciális adatsorra vonatkoztatva visszakanyarodott Newcomb általánosabban már megfogalmazott felvetéséhez: „*It is curious to remark that this law would enable us to decide whether a large collection of independent numerical results were composed of natural numbers or logarithms.*” (Newcomb, 1881). 2007-re már évente 45-50 publikáció jelent meg ebben a témában (lásd Miller, 2015; valamint https://www.researchgate.net/figure/Publications-on-Benford's-Law-from-1938-to-2007_fig2_280157559). A publikációk száma a <https://www.benfordonline.net/> oldal gyűjtése szerint az évtized második felére pedig már megközelítette az évi 150 publikációt (lásd az M1 ábrát). 2025 május közepéig ez az oldal már 2595 publikációt sorolt fel. Ezek mintegy 13 százaléka lehet olyan elemzés, amely adatsorok ellenőrzésére, csalások feltárására használja a Benford törvényt. (A publikációk 13 százalékának címében, vagy a tudományos folyóiratok nevében szerepel a „fraud” szó.) A Benford törvényről, illetve a törvényt felhasználva több magyar nyelvű tanulmány is született (Anonimus, n.d.; Gyürky & Farkas, 2015; Milicz, 2016; Tóth & Hajdu, 2017; Benkő-Kiss, 2020; Tóth, 2021). A törvényre a továbbiakban Benford-törvényként hivatkozunk.

³ E tanulmány végleges változatának írása közben újabb tanulmány jelent meg a témában (Tátrai et al., 2025), amely részben magyarázatot ad az első tanulmányokban (Tátrai & Gábos, 2005a; Élő, 2025a; Tátrai & Gábos, 2005b) feltárt anomáliákra, részben újabbakra hívja fel a figyelmet. Elsőként azt támasztja alá, hogy az anomáliák a KSH által alkalmazott hibás adatimputálásra vezethető vissza. Ennek hatása már 2014-től megfigyelhető. A KSH által alkalmazott eljárás ugyanis Magyarországon lehetetlen helyzetet eredményezett, azt, hogy a 2014-2021 között a KSH jövedelemfelvételében megfigyelt személyek 11-29 százalékának több volt a nettó, mint a bruttó jövedelme, azaz több adó- és járulékvisszatérítést kaptak, mint amennyi adót és járulékot befizettek. A Nemzeti Adó- és Vámhivatal

0.3. Az elemzés első részében röviden összefoglaljuk a Tátrai és Gábos tanulmányainak fontosabb megállapításait, majd ismertetjük a KSH-nak Tátrai és Gábos tanulmányaira reagáló sajtóközleményét (KSH, 2025).

0.4. A második részben a Benford-törvény csalásfelderítési vizsgálatokban való alkalmazásával, valamint a módszer korlátaival foglalkozunk. Röviden bemutatunk néhány példát különböző területekről (vállalati pénzügyi kimutatások, makrostatisztikai adatok, illetve parlamenti és elnökválasztások elemzéséből), amelyek szemléltetik, miként használták már ezt a módszert csalások feltárásánál.

0.5. A harmadik részben egy képzeletbeli példán keresztül mutatjuk be, hogy egy, az adatállományt csak rendkívül kis arányban érintő csalás is milyen következményekkel jár az adott adatállomány és számjegyeinek eloszlására, miközben észrevétlen marad a leíró statisztikák számára. A feltételezett csalás két módszerét is elemezzük: egy nagyon egyszerű esetet és egy olyant is, amikor a csaló valamennyire tisztában van a Benford-törvénnyel és megpróbálja kijátszani azt. Eredményül azt kapjuk, hogy még az utóbbi, szofisztikált esetben is a csalás szembevető anomáliát eredményez az adatállomány eloszlásában és jelentősen növeli az irregularitást az adatállomány utolsó számjegyeinek eloszlásában. Mindez nem jelenti azt, hogy egy igazán képzett csaló ne járhatna sikerrel: elképzelhetőnek tartjuk, hogy egy képzett csaló a Benford törvényre alapozódó csalásfelderítési módszereket is képes kijátszani, célját megvalósítani, miközben a csalás maga észrevétlen marad.

0.6. Az elemzés negyedik részében a tanulmány eredményeit foglалjuk össze és néhány következtetést fogalmazunk meg. Ezek közül a legfontosabb talán az a kézenfekvő szabály, hogy egy kutatónak, ha az elemi adatok rendelkezésére állnak, mindig ajánlott első lépésként megvizsgálni annak az változónak eloszlását, amelyet elemzése során fel kíván használni. Tátrai és Gábos (2025a; 2025b) által feltárt anomáliák és a KSH ezekre adott reakciója (KSH, 2025) nyomán, úgy tűnik, ennek a követelménynek a KSH által szolgáltatott minden adatállomány és indikátor esetében érdemes eleget tenni.

(NAV) kutatóknak adott tájékoztatása szerint ilyen esetek ténylegesen csak elvétve fordulnak elő Magyarországon. Ez a jelenség különösen a szegénységi küszöb alatti személyeknél fordult elő, az ő nettó jövedelmüket „tornázta fel” a KSH, így csökkentve a szegénységi küszöb alatt élők arányát. Tátrai et al. (2025) ezen túl a KSH jövedelemfelvételét érintő két új anomáliát is feltárt. Az első ezek közül a szegénységi rés-arány volatilitásának nagymértékű növekedése 2017-2024 között. A második pedig az, hogy 2022-es és 2023-ban publikált adatfelvételben a gyermekes és egyfős háztartások esetében a jövedelemeloszlás kétcsúcsúvá vált úgy, hogy az egyik csúcs a szegénységi küszöb alatt éppen a legszegényebbeknél látható. Mindezek újra és a korábbinál hangsúlyosabban felvetik Tátrai és Gábos eredeti tanulmányainak kérdését. *Mi is történt a KSH jövedelemfelvételének adataival 2014, illetve 2017 után?*

1. Tátrai és Gábos tanulmányának két eredménye és a KSH

1.1. Tátrai és Gábos (Tátrai & Gábos, 2005a; Élő, 2025a; Tátrai & Gábos, 2005b) két megfigyelésre hívta fel a figyelmet a Központi Statisztikai Hivatal (KSH) által szolgáltatott és az Európai Unió tagországainak jövedelmi és szociális helyzetét nyomon követő EU-SILC kutatásban megjelenő jövedelmi adatok (egy fogyasztási egységre jutó éves nettó jövedelem és egyfős háztartások éves nettó jövedelme) vizsgálata során.

1.2. Első cikkükben (Tátrai & Gábos, 2025a) ismertették az egy fogyasztási egységre jutó nettó éves jövedelem és az egyfős háztartások nettó éves jövedelmének 2005-2023 évekre vonatkozó jövedelemeloszlását mutató hisztogramokat úgy, hogy a jövedelem adatokat euróban adták meg és a hisztogramokban minden esetben csak a minta 1-90 percentilisét ábrázolták.

1.3. Az adatok elemzése alapján megállapították és részletesen bemutatták, hogy míg 2005-2017 között az egy fogyasztási egységre jutó éves nettó jövedelem és az egyfős háztartások éves nettó jövedelmi adatainak megoszlása is a jövedelemmegoszlásokra jellemző klasszikus mintázatot mutatja, addig 2018-2021 között a 100 eurós jövedelmi sávok vizsgálata alapján jellemzően a szegénységi küszöb környékén kiugróan magas esetszámok láthatók, többnyire, de nem mindig a szegénységi küszöb fölött. (Például 2020-ban a teljes mintában a küszöb alatt és fölött is van torlódás, míg 2019-ben egy 300 eurós sáv emelkedik ki a küszöb fölött).

1.4. Viszonylag kevesebbet írtak egy másik, egyébként legalább ugyanilyen fontosságú megfigyelésükről, arról, hogy 2020-ban az egyfős háztartások között 322 olyan háztartást találtak, amelyek nettó jövedelme pontosan megegyezett (3.996,3110974 euró):

„2020-ban a háztartások olyan nagyságrendű csoportosulását látjuk közvetlenül a szegénységi küszöb felett, amire nehéz magyarázatot találni. Összesen 379 olyan egyfős háztartás van, amelynek 3 900 és 4 000 euró között volt 2020-ban az éves jövedelme, ami négyszerese a következő legszámosabb esetet tartalmazó jövedelemsávnak. Elvileg elképzelhető, hogy az adatellenőrzés fázisában az egyfős háztartások egy jelentősebb részében hiányzó jövedelmet találtak, és az algoritmus szerint ezek egymáshoz nagyon hasonló jellemzőkkel bíró személyek voltak, akik emiatt hasonló vagy éppen azonos jövedelemértéket kaptak az eljárás során. Ha ez valamilyen adminisztratív módon megállapított összeg, például minimálbér, akkor elképzelhető, hogy ez okozhatta a jövedelemtorlódások egy részét. Ezzel ellentétben azonban azt találtuk, hogy az ebbe a jövedelemsávba tartozó, összesen 322 egyfős háztartás jövedelmeinek az összege fillérre pontosan megegyezik egymással (egészen pontosan 3 996,3110974 euró), ami szinte elképzelhetetlen anélkül, hogy az adatjavítás fázisában közvetlenül ne a végső

háztartásjövedelmet tartalmazó változóban javítottak vagy imputáltak volna. Még akkor is, ha feltételezzük azt a nagyon valószínűtlen helyzetet, hogy egy adott évben az egyfős háztartások tagjai közül 322-en (ami az egyfős háztartások 14%-a) ugyanakkora bért kaptak volna (pl. minimálbért), hiszen a háztartás összjövedelme más jövedelemkomponenseket is tartalmaz(hat), ami hozzáadódik a bérhez, és emiatt a számításunk alapját jelentő összes háztartásjövedelem értékeiben eltéréseket látnánk. Ráadásul a csak az egyéni keresetet tartalmazó változó eloszlásának vizsgálatakor ugyanezt a torlódást nem tapasztaltuk, tehát ezt a magyarázatot több okból is kizárhatjuk.”

[Tátrai & Gábos, 2025a]

Ez a megfigyelés azért fontos, mert – a csalásfelderítési vizsgálatok tapasztalatai erre mutatnak – az ismétlődő értékek ott, ahol nyilvánvaló adminisztratív okokból ezek nem következhetnek be, további vizsgálatra érdemes anomáliaként kezelendők.

Tátrai et al. (2025) mélyebben elemezte e háztartások adatait és a fenti megfigyelést azzal egészítette ki, hogy az euro mellett megadta ennek forintban kifejezett összegét is, ami éppen százezer forintra kerekített érték (!), 1.300.000,00 forint volt. Ez egyébként éppen a szegénységi küszöböt (1.264.430 Ft, vagyis 3887 euro) meghaladó érték. Ezen túl kimutatták, hogy ez az több száz esetben ismétlődő érték új „jött ki”, hogy a KSH változó értékű negatív jövedelemadó és járulék összegeket írt jóvá e háztartásoknál. Tehát e háztartásoknál az azonos 1,3 milliós szintre, éppen a szegénységi küszöb fölé, emelt nettó jövedelem abszurd módon meghaladta e háztartások teljes bruttó jövedelmét.

1.5. Tátrai és Gábos (2025a) tanulmányuk összefoglalójában a KSH adatok minőségével kapcsolatban az alábbi megállapításokat tették:

„Írásunkban arra törekedtünk, hogy részletesen bemutassuk a Központi Statisztikai Hivatal által az EUROSTAT számára szolgáltatott évenkénti jövedelemadatokban 2017-től megjelenő, különböző természetű adatproblémákat. Az itt feltárt rendellenességek többsége a szegénységi arányszám csökkenésének irányába mutat. (...)

Az elvárt eloszláshoz képest ilyen nagymértékű, sorozatos eltérések aggodalomra adhatnak okot, alkalmasak lehetnek arra, hogy megingassák a hitet az adatok megbízhatóságában. Ez elbizonytalaníthatja a hazai és nemzetközi adatfelhasználókat, kutatókat abban, hogy ezek az adatok alkalmasak-e valós társadalmi folyamatok leírására és magyarázatára, a politikai döntéshozókat pedig abban, hogy beavatkozásaikat valóságos vagy torzított adatokra építik-e. Ráadásul, a cikkben leírt és általunk vélelmezett imputációs eljárás sértheti az [Európai Parlament és Tanács 223/2009/EK rendeletét](#), az [Európai Parlament és a Tanács \(EU\) 2019/1700 rendeletét](#), valamint az [Európai Statisztika Gyakorlati Kódexét](#).”

A szerzők tehát “adatproblémák”-ról, “rendellenesség”-ekről és az ezekből fakadó bizalomvesztésről írnak: mivel az “ilyen nagymértékű, sorozatos eltérések

aggodalomra adhatnak okot” és “alkalmasak lehetnek arra, hogy megingassák a hitet az adatok megbízhatóságában” valamint “elbizonytalaníthatja a hazai és nemzetközi adatfelhasználókat, kutatókat ”. Ezen túl felvetik annak lehetőségét, hogy a cikkben vélelmezett imputációs eljárás “sértheti” az Európai Parlament és Tanács 223/2009/EK rendeletét, az Európai Parlament és a Tanács (EU) 2019/1700 rendeletét, valamint az Európai Statisztika Gyakorlati Kódexét.

1.6. A szerzők cikkükben nem írták, de (1) az általuk közölt ábrákból – az egy fogyasztási egységre vonatkozó éves nettó jövedelmének és az egyfős háztartások éves nettó jövedelmének 1-90 percentiliseit ábrázoló hisztogramokból – látható, hogy 2005-2023 között jövedelemnagyság szerint jelentősen széthúzódott a KSH adatfelvétel mintája: míg 2005-ben a KSH mintájának első 90 percentilise a -200 – 5.900 euró közötti sávokba esett, addig 2023-ban már a 0 – 12.300 euró közötti sávokba. Ezzel együtt (2) az időszak során, különösen 2017 után, úgy tűnik, nagyobb ugrások vannak az egyes intervallumokhoz tartozó gyakoriságok között, mint előtte. Az EU-SILC gördülő panelként működik. Így egyik évről a másikra a mintanagyság nem változik drasztikusan, de az idősor egészében jelentős különbségek figyelhetők meg a mintanagyságban (14.000-30.000 fő közé esett 2005-2023 között a felmért személyek száma). Az eloszlás alsó 90 százalék terjedelme mellett ez is szerepet játszik a 100 eurós sávokba eső esetszámok nagyságában. A minta jövedelem szerinti széthúzódásának trendje azt is eredményezheti, hogy a jövedelmi egyenlőtlenségek növekedése miatt, egy-egy 100 eurós intervallumba átlagosan kevesebb fő tartozhatott az időszak végén, mint az időszak elején. Egy ilyen trend részben indokolhatná kiugró értékek létezését. Ezt a feltételezést könnyen lehet ellenőrizni az adatfelvételek jövedelemnagyság szerinti megoszlásának vizsgálatával.

Ha létezik is ilyen, az adatfelvételek szerkezetével és nagyságával kapcsolatos jelenség, azonban kizárólag ez nem magyarázhatja Tátrai és Gábos (2025a; 2025b) által bemutatott “torlódásokat”. Ezek létrejötté ettől különböző okokra vezethető vissza.

1.7. A KSH egy sajtóközleményben reagált Tátrai és Gábos cikkeiben és a cikk eredményeit ismertető interjúban foglaltakra (KSH, 2025).

(1) Egyfelől visszautasította azt a feltételezést, hogy a KSH által publikált adatok nem lennének szakmailag kifogástalanok:

„A Központi Statisztikai Hivatal (KSH) működésében és az adatok feldolgozásában egyaránt a nemzetközi standardoknak megfelelően jár el. A szegénységi adatok forrását képező EU-SILC-adatfelvétel eredményeit az Eurostat minden évben validálja, ezt követően saját adatbázisában publikálja. A statisztikai adatok gyűjtése, feldolgozása szigorú szakmai és etikai szabályok szerint történik. A KSH minden adatsorát megfelelő szakmai felügyelet alatt dolgozza fel, és minden esetben átláthatóan, pontosan kommunikálja az eredményeket.”

(2) Másrészt olyan "vádát" utasított vissza, amelyet sem Tátrai és Gábos (2025a; 2005b) sem Élő (2025a) nem állított:

"A szándékos manipuláció vádját a KSH visszautasítja."

(3) Harmadrészt a szegénységi küszöb körüli és feletti kiugró gyakoriságokat egy módszertani váltással, valamint a minimálbérrel, a közfoglalkoztatási bérrel, valamint a garantált bérminimummal hozta összefüggésbe:

„A hivatal 2018-ban módszertani fejlesztést hajtott végre az EU-SILC-adatok feldolgozásában az adatminőség fejlesztése érdekében. Ennek részeként a jövedelmi adatok ellenőrzéséhez és a hiányzó adatok pótlásához egyedi szinten, közvetlenül a NAV és a MÁK adminisztratív kereseti adatai kerülnek felhasználásra. A jövedelem eloszlásában jelentkező kiugrások, illetve sűrűsödési pontok az adminisztratív adatforrásból származó kereseti adatoknál is megfigyelhetők a közfoglalkoztatási bér, a minimálbér, valamint a garantált bérminimum vonatkozásában.

Az EU-SILC-adatfelvétel jövedelemeloszlásában megfigyelhető kiugrás összefüggésben van a keresetek minimálbérkörnyéki torlódásával. A nettó minimálbér és a szegénységi küszöb ezekben az években legfeljebb csak kismértékben tért el egymástól, ebből következően részben magyarázható a jövedelmeknek a szegénységi küszöb körüli torlódása."

Tátrai és Gábos ezután közzétett újabb elemzése szerint azonban a KSH-nak ez az érve nem állja meg a helyét (Élő, 2025b).

(4) Negyedszer pedig a KSH Tátrai és Gábos (2025a) tanulmányát, vagy az azt ismertető cikket (Élő, 2025a) "áltudományos"-nak minősítette:

„Az áltudományos cikkben szereplő állítások ellenére a Központi Statisztikai Hivatal továbbra is hisz és elkötelezett az adatok valódi tudományos célú hasznosításának támogatásában, és változatlanul nyitott a kutatókkal folytatott szakmai párbeszédre, együttműködésre."

1.8. Nem tudjuk pontosan mi történt a KSH adataival. A KSH által adott információk szerint 2017 után került sor „*módszertani fejlesztésre*”, de Tátrai és szerzőtársai a legújabb tanulmánya szerint (Tátrai et al., 2025) anomáliák már 2014-től megfigyelhetők a KSH által közölt adatokban. Az új elemzésük már választ ad arra is, hogy milyen anomáliák állnak a szegénységi küszöb környékén korábban megfigyelt adattorlódás mögött:

„Az EU-SILC jövedelemadatai egy olyan országot tárnak elénk, ami biztosan nem lehet Magyarország: a 2014 és 2021 közötti időszakban a mintába került háztartások 4-34%-ának esetében a rendelkezésre álló éves nettó jövedelem meghaladta a bruttó háztartás-jövedelem értékét. Eszerint a személyek 11-30%-a nem fizeti az adót, hanem

kapja. Ez a magyar személyi jövedelemadó- és társadalombiztosítási járulékrendszer jellemzőit figyelembe véve lehetetlen.

(...)

Az eddig bemutatottak következtében jelentős számban fordulnak elő olyan esetek, amelyek a negatív adó- és járulékbefizetések miatt kerülnek a szegénységi küszöb fölé. Vagyis bruttó jövedelmük szegénységi küszöb alatt van, de a nettó jövedelmük szegénységi küszöb fölötti. Az ilyen módon szegénységből kilépő "küszöbátlépő" esetek aránya a teljes népességben 2014 és 2021 között 1-9%, akik tehát nagy valószínűség szerint **csak azért kerültek a statisztikában a szegénységi küszöb fölé, mert a KSH olyan adó és járulékvisszatérítéseket számolt el nekik, ami a magyarországi szabályozás szerint nem fordulhat elő.**"

[Tátrai et al. (2025), a vastagon szedett részek a szerzők kiemelései]

1.9. Nyilvánvaló, hogy egy „*módszertani fejlesztés*” meglehetősen erősen képes módosítani egy mutató értékét. Kitűnő példa erre a Világbank Doing Business országrangsorának esete. Egyes kormányok árgus szemmel nézték, hogy hol helyezkedik el országuk ezen a rangsoron és mennyit javult országuk helyezése az évek során. Narendra Modi, India miniszterelnöke például látva a javuló indiai adatokat, lelkes twitter üzenetet tett közzé 2017 október 31-én:

"Historic jump in 'Ease of Doing Business' rankings is the outcome of the all-round & multi-sectoral reform push of Team India."

[<https://x.com/narendramodi/status/925371481437044737>]

Később, elemzések kimutatták, hogy ténylegesen itt semmilyen javulásról nem volt szó: India jobb helyezése csupán a Világbank által megvalósított *módszertani fejlesztésből* következett (Sandefur & Wadhwa, 2018).

1.10. A KSH még nem adott magyarázatot arra, hogy a "*módszertani fejlesztés*" hogyan és mennyiben járult hozzá a kutatói nyilvánosság által elérhető adatoknak a Tátrai és Gábos (2025a és 2025b) által feltárt anomáliáihoz, valamint arra, hogy Tátrai et al. (2025) által bemutatott abszurd jelenségek hogyan fordulhattak elő. A KSH szempontjából öt lehetséges forgatókönyv képzelhető el – ezek közül egyik sem zárható ki előzetesen, hacsak a KSH nem mutat be olyan meggyőző bizonyítékokat, amelyek segítségével egy kivételével a többi lehetőség kizárhatóvá válik.

A. A KSH módszertani fejlesztése semmilyen kormányzati elvárással, a szegénységi küszöb alatt élők arányának kívánatos csökkentésével nincs oksági kapcsolatban. A módszertani fejlesztés nem a kormányzati elvárások teljesítése miatt történt. A KSH nem volt tudatában annak, hogy a módszertani fejlesztés egyben a kormányzati elvárás teljesülését is eredményezheti.

A1. A KSH nem tudott a jövedelemeloszlásban 2018-2022 években jelentkező kiugró gyakoriságok létezéséről, sem a 2014 óta jelentkező

abszurd jövedelmi adatokról (a megfigyelt személyek 11-30 százalékánál a bruttó jövedelem alacsonyabb a nettó jövedelemnél). Nem elemezte a módszertani fejlesztésben megvalósított adatimputálás után kapott a jövedelemeloszlás hisztogramjait és leíró statisztikáit és nem ellenőrizte azt, hogy a módszertani fejlesztés, a NAV és a MÁK adatok imputálása mennyiben befolyásolta az adatminőséget. A jövedelemeloszlásban 2018 után jelentkező anomáliákat az ezzel foglalkozó kutatók elemzéseiből (Tátrai & Gábos, 2022; Tátrai & Gábos, 2025a; Tátrai & Gábos, 2025b; Tátrai, 2025) tudta meg, addig ezek ismeretlenek voltak előtte.

A2. A KSH tudott a jövedelemmegoszlásban meglévő kiugró gyakoriságok létezéséről és a 2014 óta megfigyelhető abszurd jövedelemadatokról. Elemezte és ellenőrizte a módszertani fejlesztésben megvalósított adatimputálás után kapott jövedelemeloszlás hisztogramjait és leíró statisztikai jellemzőit és mindent rendben talált. Nem tulajdonított olyan jelentőséget a szegénységi küszöb értéke körüli kiugró gyakorisági értékeknek, amely a módszertani fejlesztés, illetve ennek egyes lépéseinek (például a hiányzó adatok imputálása) szakmai megalapozottságát, illetve az Eurostat, valamint az EU által előírt szabályok megsértését jelentette volna.

B. A KSH módszertani fejlesztése semmilyen kormányzati elvárással, a szegénységi küszöb alatt élők arányának kívánatos csökkentésével nincs oksági kapcsolatban. A módszertani fejlesztés nem a kormányzati elvárások teljesítése miatt történt. A módszertani fejlesztés azonban hozzájárult ahhoz is, hogy teljesültek kormányzati elvárások: a szegénységi küszöb alatt élők arányának csökkentése. A KSH tudott a módszertani fejlesztésnek erről a nem szándékolt mellékhatásáról.

B1 ugyanaz, mint A1.

B2 ugyanaz, mint A2.

C. A KSH módszertani fejlesztése oksági kapcsolatban volt a kormányzati elvárással, a szegénységi küszöb alatt élők arányának kívánatos csökkentésével. A módszertani fejlesztés a kormányzati elvárások teljesítése miatt történt.

Ekkor az, hogy A1 (B1) vagy A2 (B2) lehetőségek megkülönböztetése lényegtelen.

A felírt lehetőségeket az 1. táblázat foglalja össze.

1. táblázat: A lehetséges esetek: csalás, a módszertani fejlesztés anomáliái, ellenőrző vizsgálat

	Esetek	Volt-e csalás?	A KSH tudott-e a módszertani fejlesztés mellékhatásáról?	A módszertani fejlesztés után volt-e ellenőrző vizsgálat?
		(oksági kapcsolat)	(korreláció)	
1	A1	Nem volt	Nem tudott	Nem volt
2	A2	Nem volt	Nem tudott	Volt
3	B1	Nem volt	Tudott	Nem volt
4	B2	Nem volt	Tudott	Volt
5	C	Volt	-	-

1.11. A KSH sajtóközleménye (KSH, 2025) szerint a C lehetőség kizárt. Tegyük fel, hogy ez az állítás igaz. Ekkor is fontos lenne tudni azt, hogy az A1, A2, B1, vagy B2 lehetőségek közül melyik következett be. Ha az A1, akkor ez rontja a KSH hitelességét, mivel ignorált nyilvánvaló anomáliákat, illetve nem körültekintően végezte el a módszertani fejlesztés után a kapott adatok ellenőrzését; ha a A2, akkor ez is rontja a KSH hitelességét, mert nem ellenőrizte a módszertani fejlesztésnek a közzétett indikátorra gyakorolt hatásait; ha B1, vagy B2, akkor ezek is hitelességromlást jelentenek, megegyezően az A1 és A2 lehetőségekkel.

1.12. A KSH szakmai hitelességét a vonatkozó adatfelvétel kapcsán csak a szakmai nyilvánosság előtti párbeszéd, a KSH által alkalmazott *módszertani fejlesztés* minden lépéseinek feltárása és nyilvánosságra hozása, valamint minden szakmai kérdésre kiterjedő független vizsgálat és e vizsgálat eredményeinek a KSH honlapján való publikálása állíthatja helyre.

1.13. A fentiekkel összefüggésben érdemes felidézni a Világbank Doing Business rangsorával kapcsolatos szakmai kritikák és a Világbank ezekre való válaszainak történetét⁴. A rangsort először szakmai kritikák érték. Erre a Világbank először egy felületes belső vizsgálattal válaszolt, majd a megújuló kritikák hatására kénytelen volt a rangsor publikálását felfüggeszteni és egy független vizsgálatot kezdeményezni. A független vizsgálat politikai befolyásolást és ennek nyomán komoly anomáliákat tárt fel. A Világbank a független vizsgálat eredményeit tartalmazó jelentést honlapján nyilvánosságra hozta és a vizsgálati eredmények hatására a Doing Business program befejezése mellett döntött.

⁴ A Doing Business rangsort ért kritikák és a Világbanknak erre adott válaszainak történetének rövid leírását lásd az M3 mellékletben.

2. Csalásfelderítés – néhány példa

2.1. Ha nincs, és talán nem is lesz információnk a KSH által 2017 után bevezetett módszertani fejlesztés egyes lépéseiről, és a KSH nem teszi közzé olyan elemzést, amely ezek hatását vizsgálja a KSH által publikált jövedelemadatokra, akkor is elemezhető lesz az, hogy ez a módszertani fejlesztés hogyan hatott az adatfelvétel adatminőségére. Ez analóg a fekete doboz problémához: nem tudjuk milyen függvények és milyen függvény paraméterek eredményeként kapjuk azokat a kimeneti információkat, amelyeket a fekete doboz kibocsát. De ismerjük a bemenetei és kimeneti információkat és ezekből következtetni lehet a fekete doboz működésére. A kimeneti információk közül nemcsak a jövedelemeloszlását lehet elemezni, hanem segítségül hívhatók a csalásfelderítési vizsgálatok bevett módszerei is. Tátrai és Gábos (2025a) által feltárt egyik fontos anomália – bizonyos jövedelmi adatok nagyszámú ismétlődése – nyomán érdemes mélyebben is megnézni például a jövedelemadatok számjegyeinek eloszlása mennyiben felel meg a Benford-törvénynek. Ez a módszer már bevett gyakorlatnak számít statisztikai adatok ellenőrzésénél (Nigrini & Mittermaier, 1997; Dlugosz & Jann, 2009; Nigrini, 2012; Miller, 2015; Kossovsky, 2015). Egy ilyen vizsgálat más oldalról tisztázhatná, hogy vannak-e anomáliák a közzelt egyéni jövedelemadatokban, és azt is, hogy milyen eredményeket kapunk e téren a KSH módszertani fejlesztése előtti az azt követő években. Ha például az egyéni jövedelemadatok számjegyeinek eloszlásának a Benford törvény által prediktált eloszlástól való eltérését a KSH módszertani fejlesztése számottevően befolyásolná, akkor ez arra utalna, hogy a módszertani fejlesztés nyomán romlott az adatfelvétel adatminősége, mivel az adatokban korábban nem látott új irregularitás jelent meg. Azt is lehetne ellenőrizni, hogy ebből a szempontból eltérnek-e a magyar adatok más országok adataitól.

2.2. A Benford-törvény alkalmazását adathibák ellenőrzésére először Varian (1972) vetette fel gazdasági előrejelzések adatainak ellenőrzése céljából. A csalásfelderítésben az első alkalmazása vállalati pénzügyi adatok ellenőrzéséhez kötődik (Charslaw, 1988). Nemzetközi intézmények, statisztikai hivatalok által közzelt adatoknál ismerünk olyan esetet, amikor a vizsgálatok adatmanipulációra találtak bizonyítékokat. Több kutatás kimutatta, hogy a nem manipulált makroökonómiai adatok első számjegyeinek eloszlása a Benford-törvény által előírt eloszlást követ (Rauch et al., 2015). A görögországi túlzott deficit eljárással összefüggő 1999-2009 közötti makroadatokról az derült ki, hogy ezek első számjegyeinek eloszlása 2000-2002 között és 2007-2009 között számottevően eltért a Benford-törvény által meghatározott eloszlástól, míg a társadalomstatisztikára vonatkozó görög adatoknál ez a jelenség 1999 és 2009 között egyszer sem fordult elő (Rauch et al., 2015). A közgazdaság-tudományi kutatások adatainak ellenőrzésénél, az adatok kitalálása, az adatgyártás (*data fabrication*), és az adathamisítás (*data falsification*) ellenőrzésénél a Benford-törvény alkalmazása mellett érvelt Tödter (2015). Eckhartt & Ruxton (2023) a tudomány egészére nézve javasolta az eljárás alkalmazását. Véleményük szerint a tudományos publikációk által elemzett és publikációk mellett közzétett

adatok vizsgálatánál, a kutatások megbízhatóságának ellenőrzésénél és az esetleges adatcsalással kapcsolatos aggályok kivizsgálásánál érdemes alkalmazni a Benford-törvényre alapozódó vizsgálatokat.

2.3. Hozzá kell tennünk, hogy az eddigi tapasztalatok alapján a számjegyek eloszlásának vizsgálata a csalások felderítésben inkább kiegészítő bizonyítékként használható és értelmezhető. A görög makrostatisztikai adatokra vonatkozó csalás, bár *utólag kimutatható* a számjegyek vizsgálatának elemzésével (Nigrini, 2012; Rauch et al. 2015), nem a számjegyek vizsgálatának eredményei alapján került felszínre. Nyilvánvaló, hogy a tapasztalati eloszlás és a Benford-törvény által prediktált eloszlás összehasonlításánál is számolni kell mindkét fajú hiba előfordulásával. Előre nem mindig látható, hogy a vizsgált sokaság vagy minta egy tulajdonságát leíró változó számjegyeinek eloszlása „természetes esetben” Benford eloszlást követ-e⁵. Lehetséges, hogy csalásra, adatmanipulációra utaló jelet látunk ott, ahol voltaképpen ilyen nem is történt. Erre hívja fel a figyelmet Diekmann & Jann (2010) Tödter (2009) cikkével vitatkozva és óvatosságra int a Benford-törvény alkalmazásával és az így kapott eredmények interpretációjával kapcsolatban. Másrészt a számjegyek eloszlásának vizsgálata nem minden csalásra tud fényt deríteni. A Benford-törvényt alkalmazva nem mindig látunk csalásra, adatmanipulációra utaló jeleket ott, ahol ténylegesen előfordult csalás. Diekmann & Jann (2010) megkérdőjelezi a Benford-tesztek hatásosságát a tudományos csalások felderítésében, példát hozva arra, amikor a csalással létrehozott tudományos adatok megfeleltek a Benford-törvénynek. Felhívják a figyelmet arra is, hogy bár egyszerűsége miatt nagyon vonzó a számjegyek eloszlásának a Benford-törvény szerinti vizsgálata, de meglepően kevés kutatás foglalkozik ezen vizsgálatok érvényességével. Ezzel együtt lehetséges olyan körülményű vizsgálat, amely figyelembe veszi Diekmann & Jann (2010) által megfogalmazott szempontokat is (Roukema, 2015).

2.4. Kossovsky (2015) és Nigrini (2015) példák sokaságát ismerteti a Benford-törvénynek az üzleti életben elkövetett csalások feltárásánál történt alkalmazására. A <https://www.benfordonline.net/> gyűjtése több száz csalással foglalkozó tudományos közleményt sorol fel. Mi most a Nigrini által ismertetett esetek közül hármat emelünk ki. Nigrini (2015) említi Wayne James Nelson Arizona állam vagyonkezelőjének esetét, aki mintegy 2 millió dollár értékben 23 csekket állított ki egy kitalált szállítónak és az csekkeken szereplő összegeket aztán saját céljaira kívánta felhasználni. Nelsont a bíróság bűnösnek találta csalás kísérletében. A hamis csekkek számjegyei több furcsaságot mutattak. Egyrészt mind a 23 csekk a

⁵ Fewster (2009) szellemesen bemutatva, hogy a statisztikai adatok számjegyeinek eloszlása mikor és miért fogja a Newcomb-Benford törvényt követni, amellet érvel, hogy egy x statisztikai adat számjegyei Benford eloszlást mutatnak, ha x értékei több nagyságrendet fognak át és x eloszlása “kellően sima”: “We can conclude that data from any distribution will tend to be ‘Benford’, as long as the distribution spans several integers on the log10 scale—several orders of magnitude on the original scale—and as long as the distribution is reasonably smooth.” Minél szélesebb terjedelmet fog át x , számjegyeinek eloszlása annál inkább közelíti a Benford törvény által meghatározott valószínűségeket (Fewster, 2009).

100.000,00 dolláros értékhatár alatti volt. Ennek oka az volt, hogy az elkövetőnek a 100.000,00 dolláros értékhatár felett automatikus ellenőrzéssel kellett volna szembenéznie és ezért kerülte az ennél nagyobb összegeket. Másrészt a csekkeken szereplő kitalált összegek első számjegyei a Benford-törvénnyel ellentétes gyakoriságot mutattak: az első számjegyek 70 százaléka 7-es, 8-as, vagy 9-es számjegy volt. Ezen túl megfigyelhető volt az is, hogy az ez első két számjegyek között a 87-es, 88-as 93-as és a 98-as is duplán szerepelt. Az utolsó két számjegy közül pedig a 16-os, 67-es és a 83-as szerepelt duplán.

2.5. Másik eset, amit Nigrini ismertet, egy sikkasztás, amelynek elkövetője egy városi lakáshivatal biztonsági főnöke volt. Itt a sikkasztó biztonsági főnök el nem végzett munkáért állított ki csekkeket 1981 és 1991 között valóságos dolgozók neveit felhasználva, majd a csekkeket beváltotta és az összegeket saját céljaira fordította mintegy 500 ezer dollár értékben. A csekkekre írt kitalált, fabrikált összegeknél itt is megfigyelhető volt az, hogy a hamis kifizetések első számjegyeinek valószínűsége meglehetősen eltért a Benford-törvény által meghatározottól, valamint az is, hogy gyakori volt a számjegyek ismétlése is, amikor a csaló ugyanazokat az összegeket használta a hamis csekkeknél.

2.6. Az Enron könyvelési csalása az a harmadik eset, amit Nigrini elemmez. Itt a második számjegyek előfordulását vizsgálja és újra felhívja a figyelmet az értékhatárok fontosságára. Az Enron esetről a fontos gazdasági mutatóknál (nettó árbevétel, nettó nyereség, stb.) a második számjegyek gyakorisági eloszlása, és ezen belül a 0 értékek extrém magas aránya (7/12) volt árulkodó, ezek is könyvelési anomáliára utaltak. Az Enron kozmetikázta a könyvelését: az összeomlás előtti négy évben csalt a tényleges értékekhez képest: felfelé módosította az árbevételének, a nettó jövedelmének és az egy törzsrésztvényre jutó nyereségének értékeit. A második számjegyek megoszlása ezt árulta el.

2.7. A statisztikai adatok, tudományos publikációk, cégek könyvelési adatai mellett a választási csalás felderítésénél is hasznos eszköz a választási adatok számjegyeinek statisztikai elemzése (Mebane, 2010; Roukema, 2013; Roukema, 2015; Roundlett & Svolik, 2016; Skovoroda & Lankina, 2016). Mebane (2010), Roukema (2013) és Roukema (2015) az 2009-es iráni elnökválasztás 366 választási régiójának adatai alapján vizsgálta a választás első fordulójának adatainál a számjegyek gyakoriságát. Mebane (2010) a második számjegyek eloszlásánál nem mutatott ki anomáliákat, de részletesebb és a 2005-ös választási adatokat is vizsgáló elemzése már Ahmadinezsád javára történő csalásra utaló jeleket („urnatöméses csalás”, *ballot-box stuffing fraud*) talált, amit két másik jelöltnek (Karroubi és Rezaei) a modellszámítások alapján várttól elmaradó szavazatszámok tettek valószínűvé. Roukema (2013 és 2015) vizsgálata szerint az első számjegyek előfordulási gyakoriságának vizsgálata a négy elnökjelölt közül egynél (Karroubi) az *első számjegy anomáliára* utalt: a vizsgálat kimutatta, hogy a hetes számjegy gyakrabban fordult elő első számjegyként, mint ahogyan azt Benford-törvénye alapján elvárható lenne. Ezen túl, ahol a hetes volt az első számjegy, ott furcsa módon a második számjegy minden esetben nulla volt. A nulla számjegynek a vártnál gyakoribb

előfordulása is az adott választási adatok manipuláltságára, emberi beavatkozásra utalnak, lévén a nulla gyakori használata többféle pszichológiai okra vezethető vissza (Fassbender et al. 2014; Wadhwa & Zhang, 2015; Wadhwa & Zhang, 2019; Gideon et al. 2017; Isaac et al. 2021; Honda et al. 2022; Wang et al. 2025). Roukema megállapította, hogy a hetes számjeggyel kapcsolatos anomália a hat legnagyobb választási régióból háromban fordul elő, és ráadásul Mahmúd Ahmadinezsád (a későbbi nyertes) ebben a három régióban nagyobb arányban kapott szavazatokat, mint a többi választási régióban. Megfigyelte továbbá, hogy abban a 41 választási régióban, ahol Karroubira adott szavazatok száma hetes számjeggyel kezdődtek, a szavazatok száma gyakrabban volt páratlan, mint páros. Ez a jelenség e választási régiókban mind a négy jelöltre adott szavazatszámnál megfigyelhető volt. Szerinte az történhetett, hogy ezekben a választási régiókban valakik a valójában Karroubira leadott szavazatokat a későbbi győztes, Ahmandinezsád javára "írták jóvá". Arra a következtetésre jutott, hogy ez az anomália körülbelül egymillió szavazatot érinthet. Az eredmények értelmezésénél Roukema arra is felhívta a figyelmet, hogy a Benford törvényt alkalmazva nem biztos, hogy minden anomáliát meg lehet találni a választási eredményekben – vagyis a csalás mértéke ennél jelentősebb is lehetett. Az elvégzett elemzés Roukema szerint csak azt jelzi, hogy az általa feltárt anomáliát valahogy nem sikerült elrejteni.

2.8. Rundlett és Svolik (2016) többek között a 2011-2012-es orosz parlamenti választások és elnökválasztás adatait vizsgálta. Az utolsó számjegyek kerekítése és a szavazati arányok közötti összefüggést figyelt meg: a szavazati arányoknak az öt többszöröseire való kerekítése gyakoribb volt azokban a választási körzetekben, amelyekben Putyint (vagy pártját) "hozták ki" győztesnek a választási bizottságok.

2.9. Skovoroda és Lankina (2016) a 2012-es orosz elnökválasztás adatait vizsgálta. Az utolsó számjegyek gyakoriságának regionális különbségeit elemezték. Azt találták, hogy az utolsó számjegyek megoszlásának vizsgálata alapján feltárt választási csalás mértéke és előfordulása szignifikánsan eltért választási körzetek szerint. A csalás azokban a régiókban volt erősebb és gyakoribb, ahol hagyományosan erős a "Kreml tisztelet" (az észak-kaukázusi köztársaságokban). Szoros összefüggést találtak regionális szinten az utolsó számjegy anomáliák és a hamis választási részvételi adatok előfordulása között. A vizsgálat kimutatta, hogy ahol megfigyelhető volt az utolsó szám anomália, ott aránytalanul nagy volt a Putyinra voksolók száma és ahol előfordult szavazatok ilyen "újraelosztása", ott gyakoribb volt a szavazatok téves számolása is. Továbbá a független megfigyelők által jelentett különböző fajta csalások regionális előfordulása szoros összefüggésben volt az utolsó számjegy anomália előfordulásával. Ez a tény egyrészt az utolsó számjegyek elemzésének hasznosságát húzza alá, másrészt azt, hogy a helyi választási bizottságok különböző csalási stratégiákat követtek a kívánt végeredmény (a Putyinra adott szavazatok "növelése") elérésére.

3. Közbeszerzések – egy feltételezett csalás

3.1. A következőkben egy hipotetikus példán elemezzük, hogy egy csalás milyen eredményekkel járna egy statisztikai változó eloszlására és e változó utolsó számjegyeinek eloszlására. Ehhez feltesszük, hogy létezik egy X elemző, aki a feltételezett csalást elköveti: a közbeszerzések egy jellemzőjét, a szerződések nettó értékét hamisítja bizonyos esetekben. Az a célja, hogy bizonyos cégek által nyert és adott értékhatár alatti szerződések értékét „feltornássa” az adott értékhatár fölé. Itt tehát nem adatgyártásról, adatok kitalálásáról, hanem egyszerű adathamisításról lesz szó. Az elemzés célja annak bemutatása, hogy a csalás mennyiben fogja változtatni azon jellemző megoszlását és alapstatisztikáit, valamint utolsó számjegyeinek eloszlását, amelynél elkövették és a csalás mennyiben járul hozzá a csaló céljának eléréséhez.

3.2. Az Orbán rendszer elemzésének egyik központi kérdése, hogy a rendszer-közelben cégek milyen súlyt képviselnek a magyar gazdaságban. Mekkora e cégek részesedése a GDP-ben, a foglalkoztatásban, vagy akár a nemzeti vagyon mekkora része van a rendszerhez köthető szereplők („vállalkozók”) kezében?⁶

Ezekre a kérdésekre még nem tudunk válaszolni, de egy olyan kérdésre igen, hogy mekkora volt 2011-2023 között az Orbánhoz, illetve kormánya tagjaihoz közvetlenül kapcsolódó *top szereplőkhöz*⁷ köthető cégek részesedése a magyarországi közbeszerzések piacán.

3.3. Tételizzük fel, hogy egy képzeletbeli X kutató tanulmányt készít erről a kérdéssel, és ebben úgy manipulálja az adatokat, hogy egy általa előre kiválasztott határérték alatt csökkenjenek az Orbán rendszer *top szereplői* és egyéb piaci szereplők cégei által megnyert közbeszerzései szerződések, illetve szerződésrészek⁸ szerződéses érték szerinti megoszlásában megmutatkozó különbségek. Ebből a célból úgy hamisítja az adatokat, hogy egy értékhatár alatt az egyéb magyar cégek által nyert szerződések értékeit a szóban forgó értékhatár fölé tornássa, így csökkentve ezen értékhatár alatt az Orbán rendszer *top szereplőin* kívüli vállalkozások (egyéb cégek) arányát. A feltételezett csaló X kutató célja az, hogy kimutassa, hogy 2011-

⁶ A járadékvadászatra épülő rendszer egyik alapkérdése, hogy a járadékvadászatra alapuló szektor mekkora szeletet tudott leszakítani a szabadpiaci tortából, mennyire sikeresen rombolta le a szabadpiaci szektort (Murphy et al., 1993).

⁷ Ehhez Tóth és Hajdu (2022) által alkalmazott listát kiegészítettük egy újabb szereplővel (Jászai Gellért). A rendszer *top szereplőinek* listája így tizenhárom nevet tartalmaz, akikhez mindössze 44 olyan cég köthető, amelyek legalább egy közbeszerzési szerződést nyertek egyedül vagy konzorciumban 1999 óta.

⁸ A Közbeszerzési Értesítő a közbeszerzések legfontosabb adatait (nyertes neve, pályázatok száma, pályázók, nyertes összeg) legrészletesebben szerződésrészek szintjén tartja nyilván. Egy közbeszerzési szerződés vagy egy, vagy több részből áll. Ezért érdemes az elemzést szerződés szinten végezni, ha a szerződés egy részből áll, és szerződés részek szintjén, ha a szerződés több részből áll. A továbbiakban a szerződéseket, ahol egy vagy több szerződésrész volt, az egyszerűség kedvéért szerződésnek hívjuk.

2023 között az Orbán rendszerben a rendszer top szereplői és az egyéb magyar cégek is hasonló arányban nyernek nagy értékű közbeszerzéseket. Nem nagyon lehet tehát beszélni politikai favoritizmusról.

2. táblázat: Az Orbán rendszer top szereplőihöz, köztük Mészáros Lőrínchez köthető cégek részesedése a magyar közbeszerzések összes nettó értékéből 2011-2023, %

	Mészáros	Top szereplők
2011	0,1	4,9
2012	0,1	6,2
2013	0,2	12,8
2014	1,0	8,0
2015	0,3	2,1
2016	3,4	16,4
2017	7,2	25,8
2018	4,6	16,2
2019	7,9	20,9
2020	9,9	25,0
2021	15,3	22,5
2022	12,1	19,8
2023 ^a	3,7	8,5
Összesen	6,2	15,5

Megjegyzés: keretszerződésekkel együtt; a: előzetes adatok

Forrás: CRCB

3.4. Az Orbán rendszer top szereplői által nyert nettó összeg arányát úgy kapjuk, hogy a hozzájuk köthető cégek által egyedül, vagy konzorciumban nyert közbeszerzési szerződések vagy szerződésrészek nettó értékét elosztjuk az összes közbeszerzési szerződés és szerződésrész nettó értékével⁹. Ezt az arányt kiszámoltuk a 2011-2023 évekre és az egész időszakra vonatkozóan (2. táblázat)¹⁰. Látható, hogy önmagában Mészáros Lőrínchez köthető (formálisan az ő tulajdonában lévő) cégek az összes közbeszerzési érték közel 6 százalékából részesedtek 2011-2023 között. Arányuk 2021-ben meghaladta a 15 százalékot, 2022-ben pedig a 12 százalékot. A top szereplőkhöz kapcsolódó 44 cég részesedése pedig az egész vizsgált időszak alatt

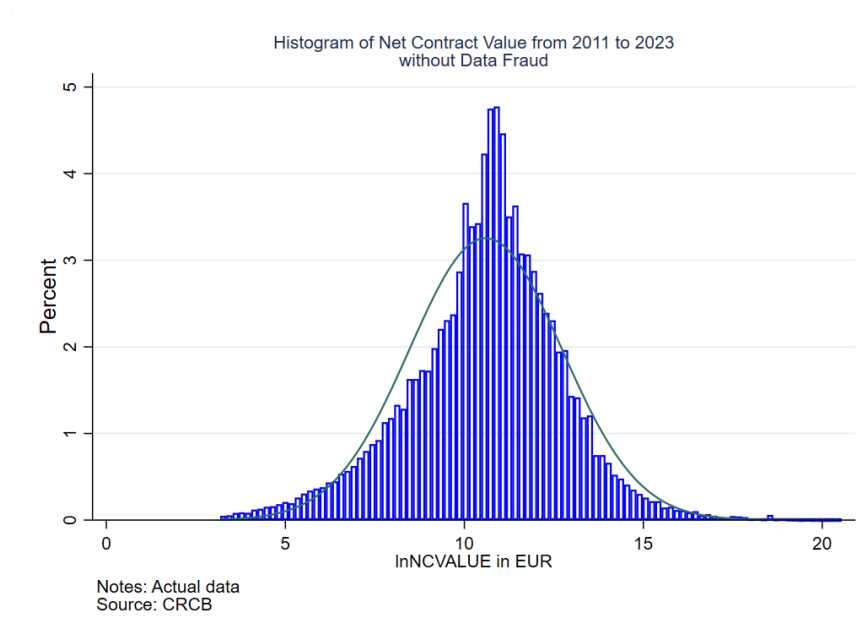
⁹ A félreértések elkerülése végett érdemes kihangsúlyozni, hogy itt nem arról van szó, hogy a top szereplők mekkora részt hasítanak ki a közbeszerzési szerződések megnyerésével a közbeszerzési piacon elérhető összes nyereségből. Ezt csak úgy lehetne kiszámolni, ha tudnánk azt, hogy az adott tranzakcióból (közbeszerzési szerződésből) származó összes nyereség hogyan oszlik meg a nyertes konzorciumi partnerek között, valamint azt, hogy a top szereplők által nyert szerződéseknél az összes nyereségben mekkora arányt képvisel a korrupciós járadék és mekkora arányt a szokásos piaci profit.

¹⁰ A vizsgált időszakra vonatkozó nettó szerződéses értéket euróra számítottuk át, és az egyszerűség kedvéért az 408,29 HUF/EUR árfolyammal számoltunk. A számítások során figyelembe vettük a keretszerződéseket és kizártuk a 10.000 forint alatti értékeket.

meghaladta a 15 százalékot. Hozzá kell tennünk, hogy ebben az időszakban több ezer cég nyert közbeszerzést Magyarországon: a fennmaradó 85 százalék e több ezer cég között oszlik meg.

3.5. A szerződések hisztogramját a nettó szerződéses érték logaritmusára szerint az 1. ábrán láthatjuk. Ez meglehetősen jól közelíti a normális eloszlást.

1. ábra: A nettó szerződéses érték logaritmusának hisztogramja, 2011-2023



Megjegyzés: lnNCVALUE: nettó szerződéses érték logaritmusára; keretszerződésekkel együtt.

3.6. Az Orbán rendszer top szereplői inkább nagy volumenű szerződéseket nyertek, mint a rajtuk kívüli cégek. Az általuk nyert nettó összegek átlaga és mediánja számottevően magasabb az egyéb magyar cégre jellemző megfelelő értékektől (lásd a 3. táblázatot és a 2. ábrát). A top szereplők a kisebb összegű szerződésekért már nem annyira hajoltak le. Az Orbán rendszer top szereplői esetében az egyéb magyar cégek középként megfigyelhető kiugrás tizenhat közbeszerzési szerződést takar, amelyeknél a nyertes ár minden esetben 24 millió forint volt. Ez a jelenség feltehetően összefüggésben van azzal, hogy a korrumpált szereplők számára egy időben kifizetődő volt a szerződéses értéket közvetlenül a 25 millió forintos értékhatár alá szorítani (Tóth & Hajdu, 2017). E szerződések nemcsak a szerződések nyertesei, de a megrendelők (ajánlatkérők) is különbözőek voltak. E szerződések megkötésére 2013 és 2023 között került sor. Háromat 2015-ben, négyet 2016-ban, kettőt 2013-ban, a fennmaradó években pedig egyet-egyet kötöttek meg. a csalásfelderítési vizsgálatok tapasztalatai alapján a szerződéses értékek ilyen nagyszámú ismétlődése anomáliának tekinthető.

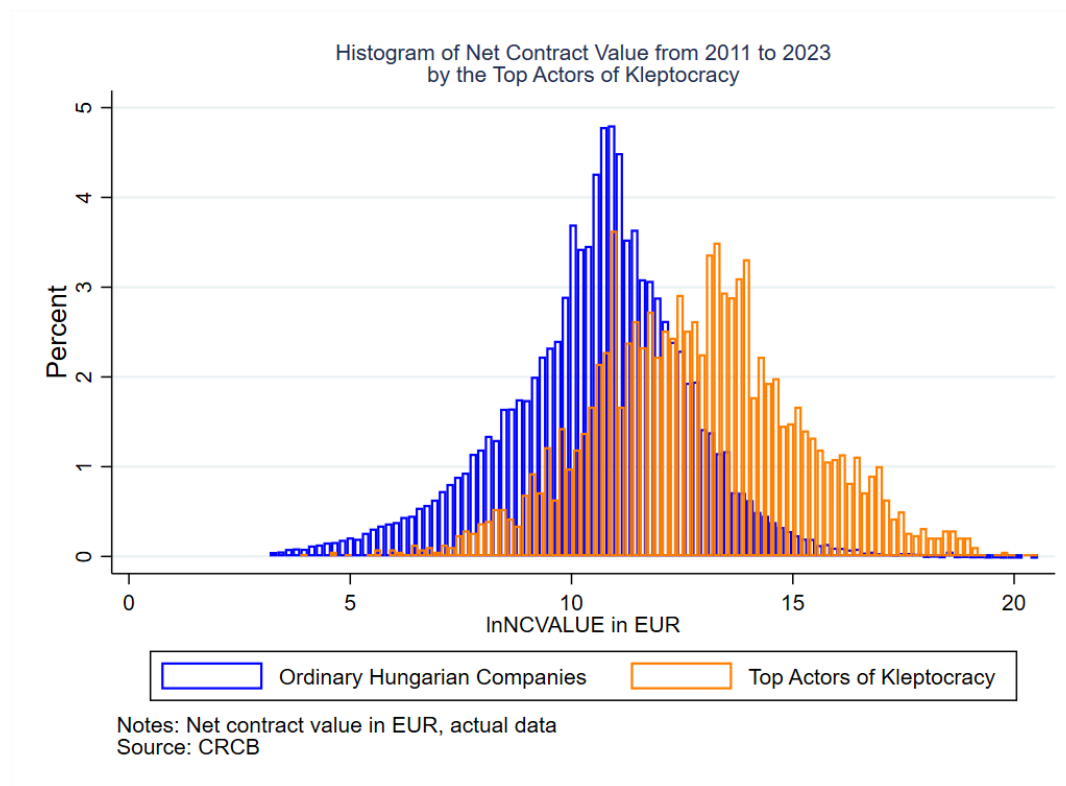
3. táblázat: Az Orbán rendszer top szereplői és egyéb magyar cégek által nyert szerződések nettó értékének alapstatisztikái 2011-2023, millió EUR

	Medián	Átlag	SE	N
Top szereplők	0,379	5,432	0,443	3.774
Egyéb cégek	0,046	0,472	0,011	237.307
Összes szerződés	0,047	0,549	0,013	241.081

Megjegyzés: keretszerződésekkel együtt.

Forrás: CRCB

2. Ábra: A nettó szerződéses érték logaritmusának hisztogramja az Orbán rendszer top szereplői és az egyéb magyar cégek által nyert szerződéseknel, 2011-2023



Megjegyzés: *lnNCVALUE*: nettó szerződéses érték logaritmus, euróban számolt érték; keretszerződésekkel együtt.

3.7. A képzeletbeli *X* kutató a feltételezett adatmanipulációnál úgy járt el, hogy nem nyúlt a top szereplők adataihoz, nehogy a csalás feltűnő legyen, hanem a top-szereplőkön kívüli cégek (egyéb magyar cégek) által nyert alacsony szerződés értékeket „javította fel” és módosította egy értékhatár fölé. Legyen ez az értékhatár (v^*) 1100 euró (408,29 HUF/EUR árfolyammal számolva 449.119 forint). *X* először úgy számolt, hogy a tényleges szerződéses értéket (v) minden $v < v^*$ esetben egyszerűen egy hamis szerződéses értékkel (v') helyettesítette, ami egy euróval

nagyobb volt, mint v^* ($v' = v^* + 1$). Ezt a „transzformációt” először, hogy jobban rejtve maradjon a csalás, csak akkor végezte el, ha a nyertes cég nem top szereplő cége volt, ha a szerződést 2014 után nyerték meg és a szerződésért több mint egy ajánlat versengett (a korrupciós kockázat alacsony volt). Ekkor a hamisított érték logaritmusát (Inv'_1) az alábbiak szerint határozta meg:

$$Inv'_1 = 7,004 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } Inv<7,004$$

ahol cr $[0,1]$ a korrupciós kockázat változója, amelynek értéke 0, ha több mint egy ajánlat versenyzett a szerződésért, és 1, ha csak egy ajánlat volt; a $year$ a szerződés megkötésének éve; a top_crony $[0,1]$ azt mutatja, hogy a nyertes cég a top szereplők egyikéhez tartozott-e vagy sem. Ennek értéke 1, ha a nyertesek között volt egy top szereplőhöz tartozó legalább egy cég és 0, ha egyéb cég, vagy cégek konzorciuma volt a nyertes.

Később, második nekifutásra, miután megismerte a Benford-törvénynek a számjegyek eloszlására vonatkozó predikcióját, hogy a csalás még kevésbé tűnjön fel, X már egy sokkal szofisztikáltabb eljárást alkalmazott és a hamis szerződéses értékek logaritmusát (Inv'_2) az alábbi függvény alapján határozta meg:

$$Inv'_2 = 7,01 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=1 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,02 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=2 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,03 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=3 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,04 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=4 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,05 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=5 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,06 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=6 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,07 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=7 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,08 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=8 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,09 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=9 \text{ \& } Inv<7,004$$

$$Inv'_2 = 7,10 \text{ ha } cr=0 \text{ \& } year>2014 \text{ \& } top_crony=0 \text{ \& } l3dignity=9 \text{ \& } Inv<7,004$$

ahol $l3dignity$ $[0,1,...,9]$ azt mutatja, hogy a szerződés nettó értékénél a százás helyiértéken lévő számjegyek között melyik volt.

X azt feltételezte, hogy ezzel az eljárással a Benford-törvényt alkalmazó ellenőrzések elől is el tudta rejteni a csalást.

3.8. A két eljárással csak minimális mértékben történt adathamisítás 2015 és 2023 között. Az X az adott időszakban az összes szerződés mindössze 3,2 százalékánál hamisította meg a közbeszerzési szerződések értékének adatait (lásd a 4. táblázatot).

4. táblázat: A hamisított szerződések száma és aránya a közbeszerzési szerződések között, 2015-2023

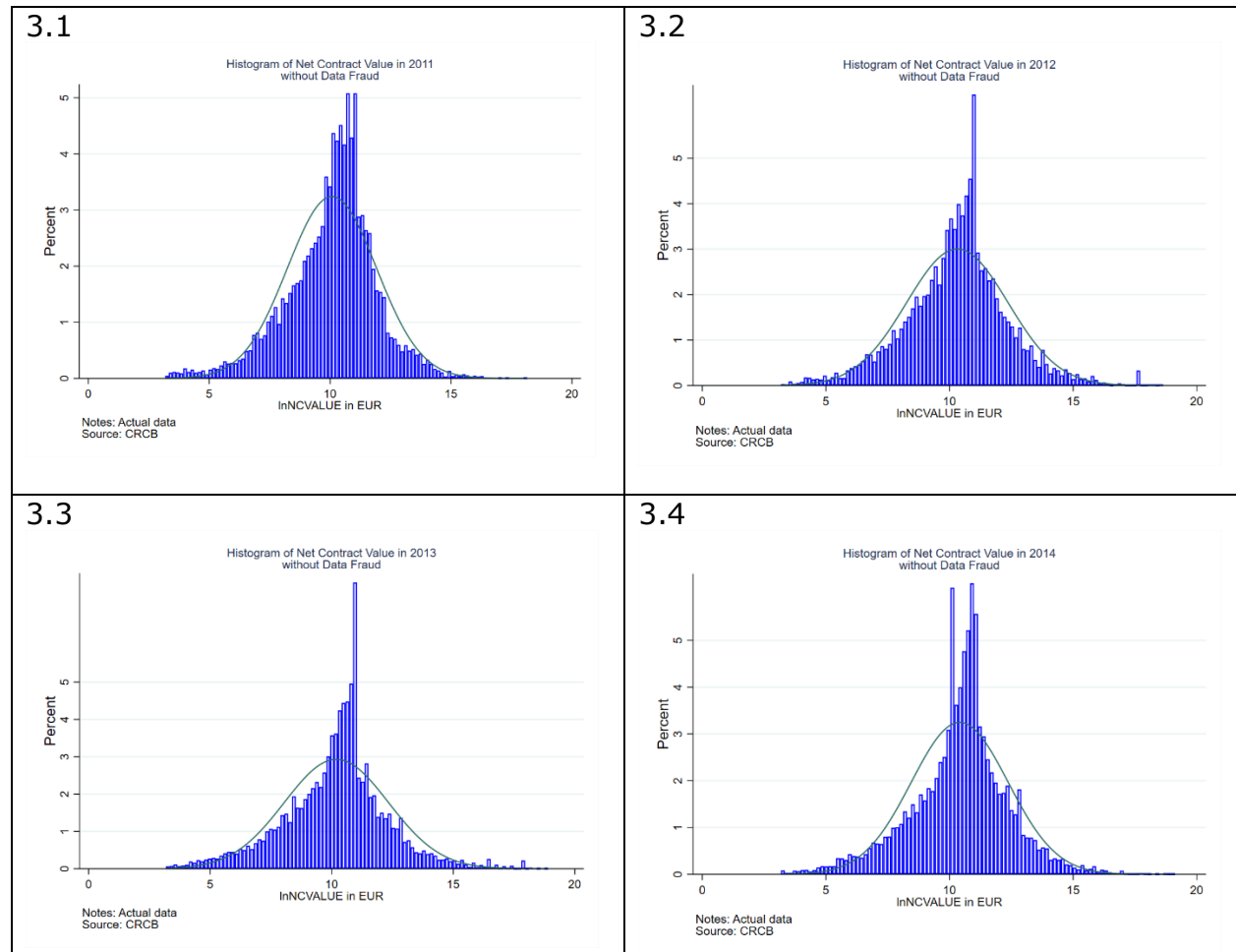
Évek	Szerződések, száma ahol nincs hamisítás	Szerződések száma, ahol van hamisítás	Összes szerződés száma	Hamisított szerződések aránya, %
2015	22306	736	23042	3,2
2016	16264	798	17062	4,7
2017	16595	693	17288	4,0
2018	22156	651	22807	2,9
2019	18630	700	19330	3,6
2020	16008	583	16591	3,5
2021	17658	385	18043	2,1
2022	17241	414	17655	2,3
2023	16484	371	16855	2,2
Összesen	163342	5331	168673	3,2

Megjegyzés: egyszerű hamisítási módszer; keretszerződésekkel együtt.

Forrás: CRCB

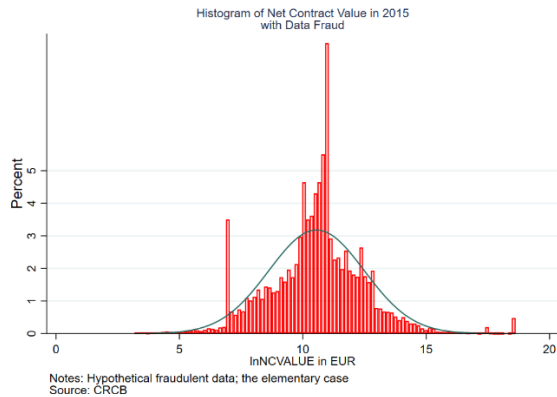
3.9. A tényleges és az első nekifutásra hamisított nettó szerződéses érték logaritmusának ($\ln v$ és $\ln v'_1$) hisztogramjait az egyes években a 3.1-3.13 ábrák mutatják. Ezekből jól látszik, hogy 2011 és 2014 között, amikor nem történt adathamisítás, a nettó szerződéses érték logaritmusának hisztogramja közel áll a normál eloszláshoz. Ellenben azokban az években (2015-2023 között), amelyeknél történt adathamisítás, a csaló X kutató által fontosnak tartott határérték feletti értéknél a szerződések száma természetellenesen torlódik: rendkívüli mértékben megnő, majd jelentősen csökken az adott értékközbe eső szerződések száma. Ez nyilvánvaló, mivel így definiálta a csalást: a célja az volt, hogy a szerződések értékét az adott értékhatár „fölé nyomja”.

3.1-3.14. ábra: A nettó szerződéses érték (Inv és Inv') hisztogramjai 2011-2023

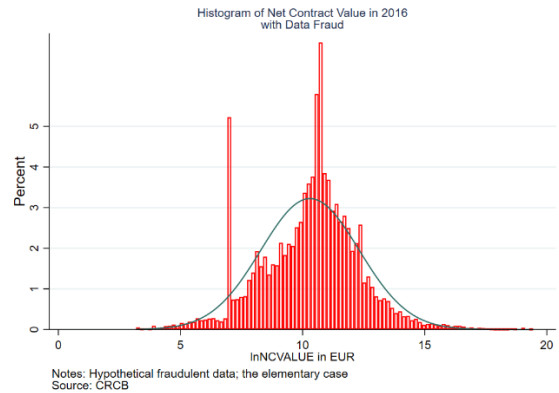


Megjegyzés: $lnNCVALUE$: nettó szerződéses érték logaritmus, ami az eredeti adatok esetében v , az egyszerű csalásnál v' logaritmusával egyenlő; keretszerződésekkel együtt; euróban számolt értékek.

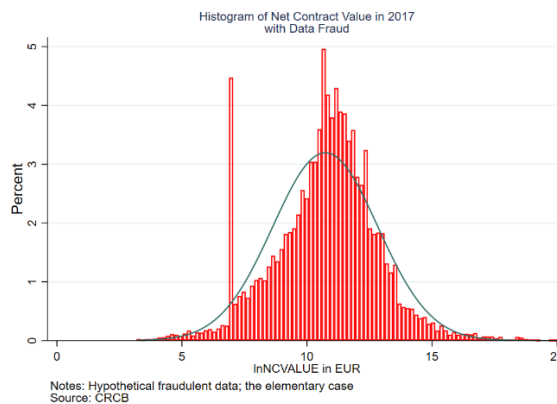
3.5



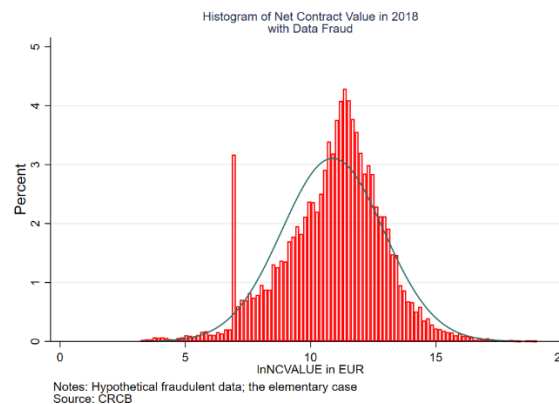
3.6



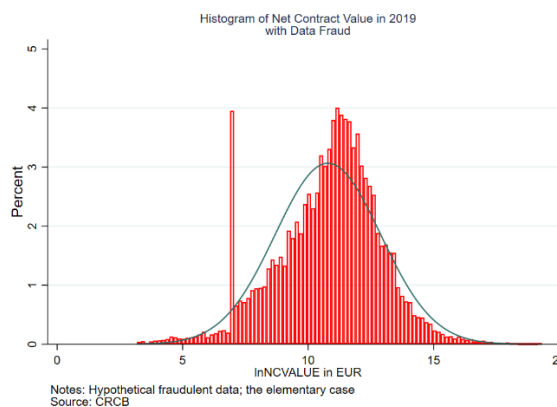
3.7



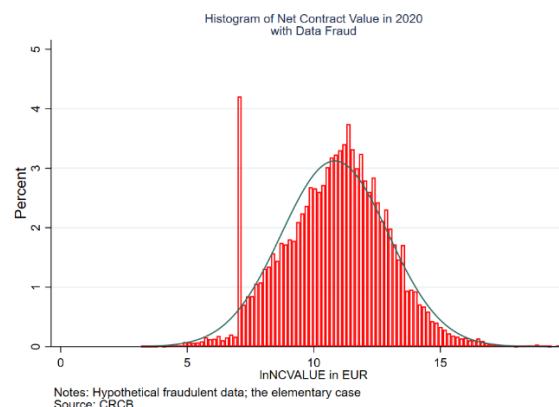
3.8



3.9

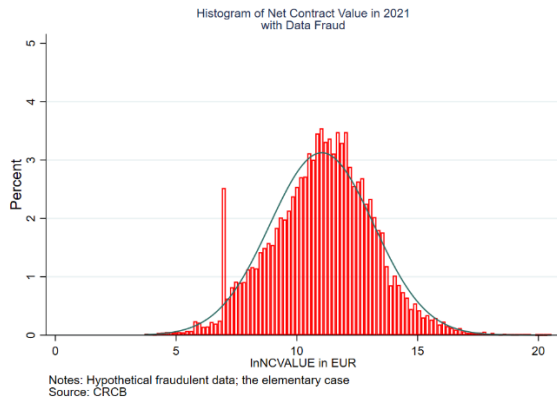


3.10

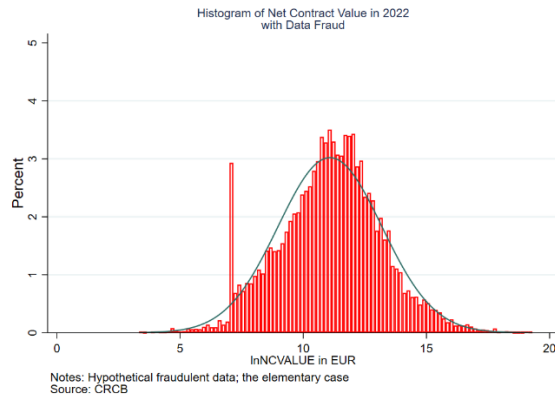


Megjegyzés: $\ln NCVALUE$: nettó szerződéses érték logaritmus, ami az eredeti adatok esetében v , az egyszerű csalásnál v_1 logaritmusával egyenlő; keretszerződésekkel együtt; euróban számolt értékek.

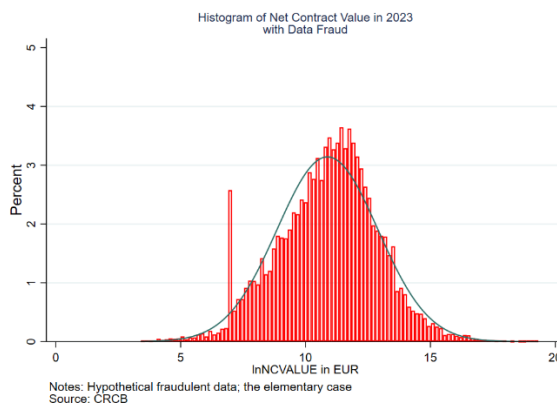
3.11



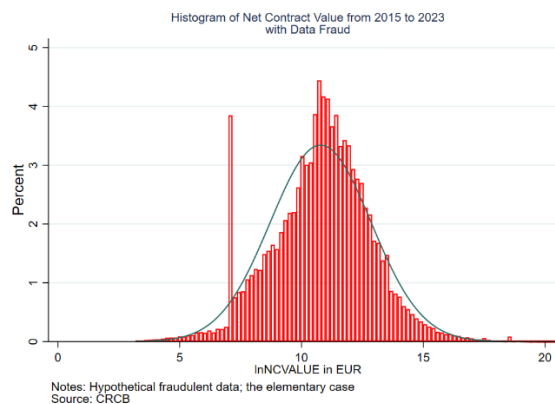
3.12



3.13



3.14



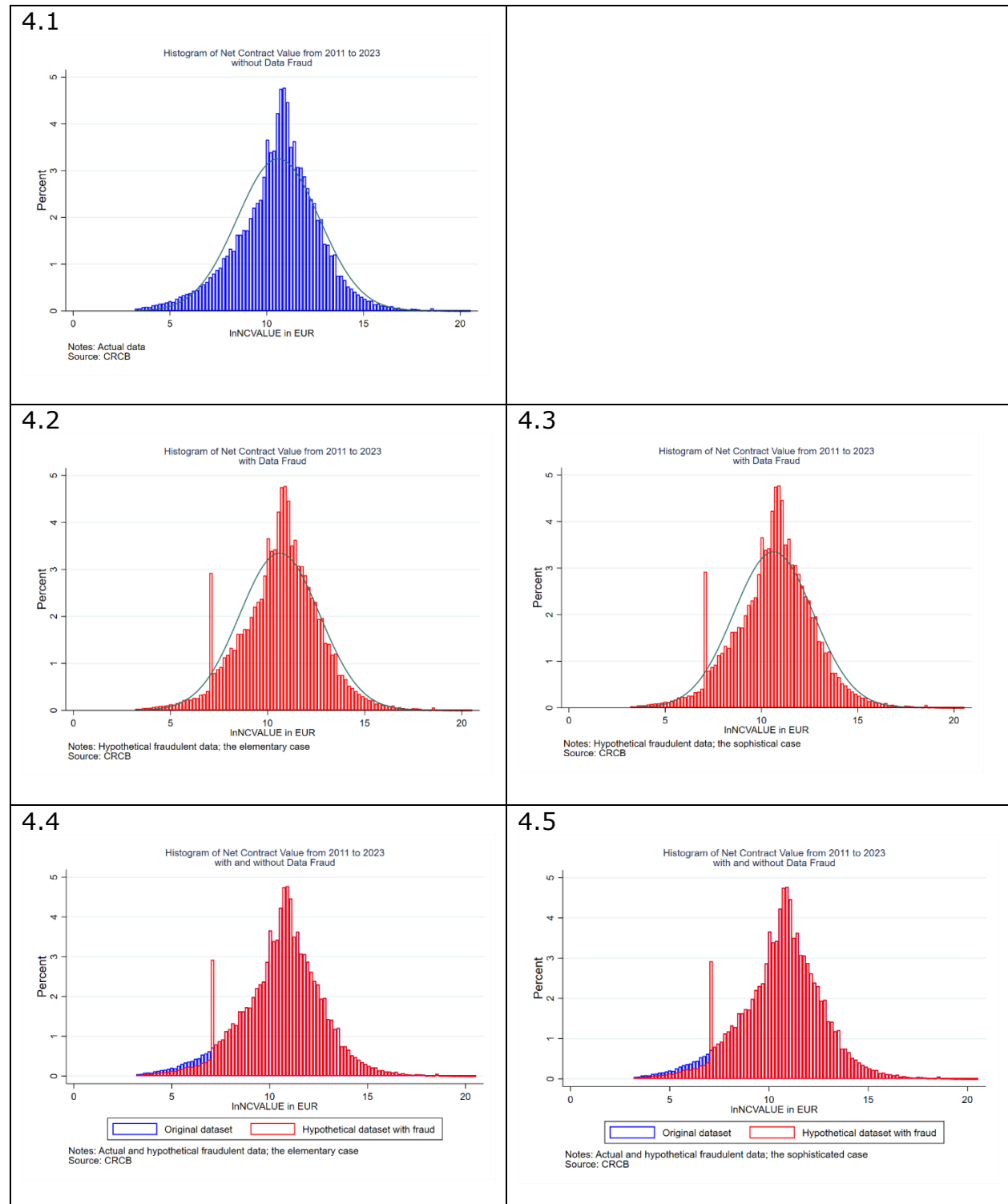
Megjegyzés: lnNCVALUE: nettó szerződéses érték logaritmus, ami az eredeti adatok esetében v , az egyszerű csalásnál v_1 logaritmusával egyenlő; keretszerződésekkel együtt; euróban számolt értékek.

A hamis adatokat tartalmazó, egész időszakra vonatkozó hisztogramon (3.14 ábra) is jól látszik ez a jelenség.

3.10. Ha egymás után, majd egyszerre ábrázoljuk a tényleges szerződéses adatokra, az egyszerű csalásra (lnv_1), majd a szofisztikált csalásra (lnv_2) vonatkozó és a 2011-2023 közötti adatokat tartalmazó hisztogramokat, akkor szemléletesebben látjuk az eredményt (lásd a 4.1-4.4 ábrát).

3.11. Az adathamisítással X célja az volt, hogy csökkentse az egyéb magyar cégeknél a kis volumenű szerződések arányát. Ez természetesen sikerült: több mint két százalékponttal csökkent ez az arány (lásd az 5. táblázatot). De igaz az is, hogy még a csalás után is nagy különbség van a top szereplők és az egyéb magyar cégek között. A csalás után az előbbieknél a szerződések 0,8 százaléka esik a meghatározott értékhatár (1100 euro) alá, míg az utóbbiaknál a szerződések 4,8 százaléka esik ide. A nyertes cégek két csoportjánál a szerződések nagyság szerinti összetétele továbbra is meglehetősen eltér (lásd az 5. ábrát).

4.1-4.5 ábra: A nettó szerződéses érték logaritmusának eredeti és hamis adatokat tartalmazó hisztogramjai, 2011-2023



Megjegyzés: $\ln NCVALUE$: nettó szerződéses érték logaritmus, ami az eredeti adatok esetében v , az egyszerű csalásnál v_1 , a szofisztikált csalásnál és v_2 logaritmusával egyenlő; keretszerződésekkel együtt; euróban számolt értékek.

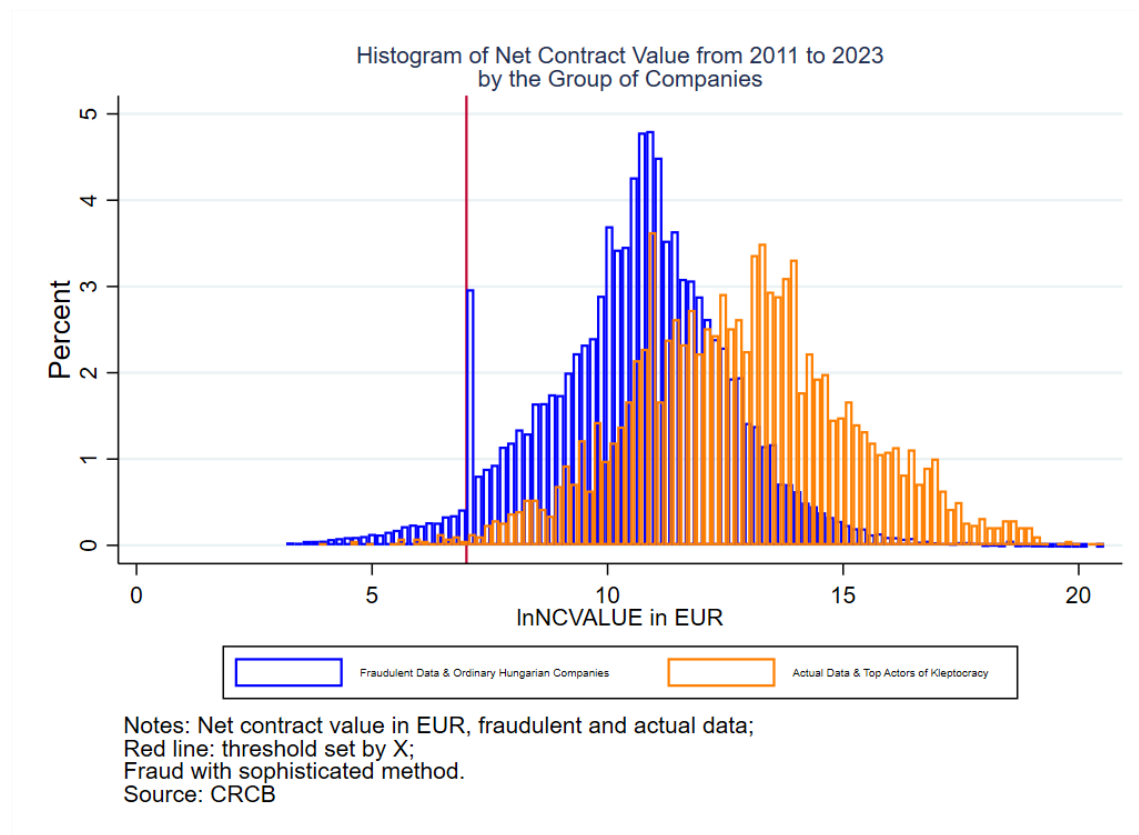
5. táblázat: A 1100 eurónál alacsonyabb összegű szerződések aránya, %

	Eredeti adatok (v)	Hamisított adatok, egyszerű eljárás (v'_1)	Hamisított adatok, szofisztikált eljárás (v'_2)
Az Orbán rendszer top szereplői	0,8	0,8	0,8
Egyszerű magyar cégek	7,0	4,8	4,8

Megjegyzés: keretszerződésekkel együtt.

Forrás: CRCB

5. ábra: A nettó szerződéses érték logaritmusának hisztogramja az Orbán rendszer top szereplői és az egyéb magyar cégek által nyert szerződésekénél a csalás után, 2011-2023



3.12. Az eredeti és a hamisított adatoknál a lehetséges anomáliák ellenőrzésére alkalmazzuk a csalásfelderítésben jól ismert és használt utolsó számjegy tesztet (*last digit test*)! Ez azt vizsgálja, hogy egy változónak egy adott populációban felvett értékeinél a változó utolsó számjegyeinek eloszlása különbözik-e számottevően az uniform eloszlástól, vagy sem. A nullhipotézis az, hogy nettó szerződéses érték utolsó számjegyeinek eloszlása nem tér el szignifikánsan a várt (uniform) eloszlástól. Ezt χ^2 próbával ellenőrizzük ($df=9$ és $p=0,05$ valószínűségi szint mellett). Az

eredmények szerint már az eredeti szerződéses adatok utolsó számjegyének eloszlása is számottevően eltér az uniform eloszlástól (lásd a 6. táblázatot). Ezt nem magyarázza az, hogy a szerződések forintban számított értékei között viszonylag nagy valószínűséggel vannak ezerre, tízezerre, vagy százezerre kerekített értékek (lásd a 7. táblázatot), ugyanis itt euróba átszámolt szerződéses értékekkel számoltunk¹¹.

6. táblázat: Az utolsó számjegy teszt a szerződések nettó értékénél, 2011-2023

	Eredeti adatok (v)	Hamisított adatok, egyszerű eljárás (v'_1)	Hamisított adatok, szofisztikált eljárás (v'_2)
Khi ²	51,330	1081,176	209,742
Cramer V	0,015	0,067	0,029

Megjegyzés: keretszerződésekkel együtt; kritikus érték = 16,919 ($df = 9$, $p=0,05$); euróban számolt értékek.

Forrás: CRCB

7. táblázat: Kerekített értékek aránya a szerződések nettó értékén belül (10^3 , 10^4 , 10^5 és 10^6 kerekítések) 2011-2023, %

	%
Ezer forintra kerekített értékek	40,7
Tízezer forintra kerekített értékek	28,6
Százezer forintra kerekített értékek	17,9
Egymillió forintra kerekített értékek	9,8

Megjegyzés: keretszerződésekkel együtt, forintban számolt értékek.

Forrás: CRCB

3.13. A 6. táblázatból az is látszik, hogy az adathamisítás újabb irregularitást visz az adatokba: az eredeti adatokra vonatkozó khi² értékhez képest (51,3) az egyszerű eljárással hamisított adatoknál a khi² értéke közel hússzorosára nő (1081,2), a szofisztikált eljárás esetén pedig négyszeresére (209,7). A bemutatott módszerekkel elkövetett csalás tehát nemcsak a hisztogramokon látszik, hanem az utolsó számjegyek eloszlását is kimutathatóan módosítja.

3.14. Az évente számolt khi² értékek jól mutatják (lásd az 6. ábrát) egyrészt, hogy egy ilyen minimális arányú csalás, ami az összes szerződés alig több mint 3 százalékát érinti, milyen nagymértékű anomáliát tud eredményezni az utolsó számjegyek eloszlásában. Még a szofisztikált módszer estében is szembetűnő a különbség az eredeti adatokhoz képest. Másrészt az is látszik, hogy a kimutatott

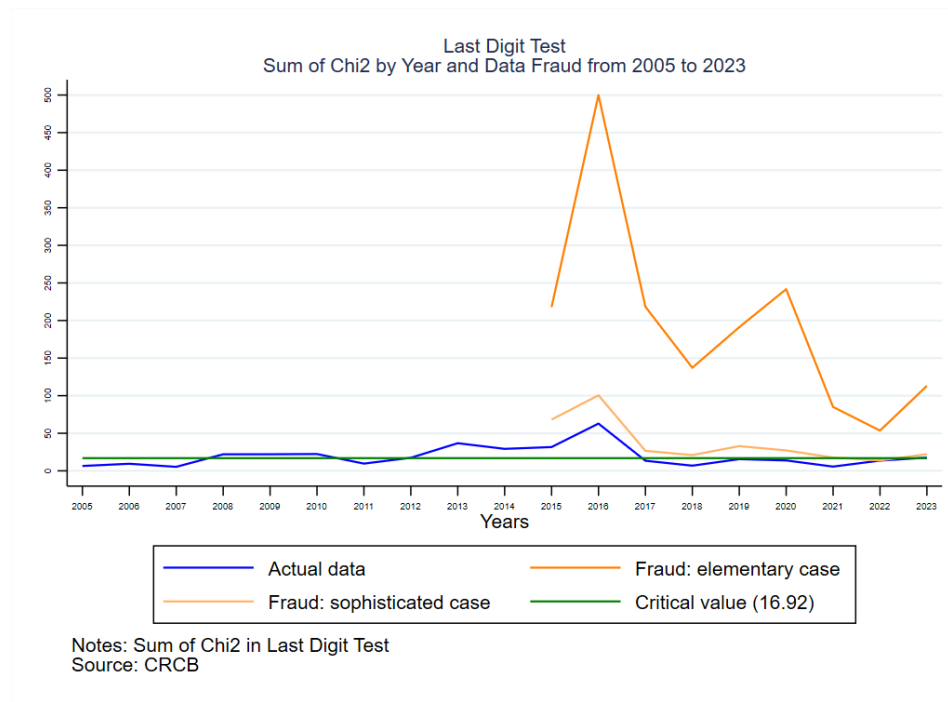
¹¹ Az 1000 forintra kerekített forintban számolt szerződéses értékek euróba átszámolt értékei utolsó számjegyeinek eloszlása nem tér el számottevően a Benford törvény által meghatározott eloszlástól.

anomália mértéke összefüggésben van azzal is, hogy a szerződések mekkora arányát érinti a csalás: amikor a szerződések kisebb arányát érinti, ott alacsonyabb khi^2 értékeket kapunk. Miközben a csalás által érintett szerződések aránya csökkenő (4,7 százalék 2016-ban és 2,2 százalék 2023-ban), a khi^2 értékek is csökkenő tendenciát mutatnak.

3.15. A 6. ábrán az eredeti adatok (kék vonal) a kritikus érték (zöld vonal) közelében ingadoznak. A 2013 és 2016 közötti kiugrás valószínűleg egy másik értékhatár anomáliával van összefüggésben: az akkor hatályos közbeszerzési törvények a 25 milliós értékhatár felett nyílt pályáztatást írtak elő. Ekkor a szereplők (ajánlatkérők és versenyző cégek) abban voltak érdekeltek, hogy az eredeti projektet akár két részre bontva, két beszerzéssé alakítsák át és így az egyes részeket a 25 milliós értékhatár alá nyomják. Így már meghívásos eljárást is választhattak, ami gyorsabb és korrupt szereplők esetén a korrupciós tranzakciót is könnyebben megvalósító megoldás volt (Tóth & Hajdu, 2017).

3.16. A fentiek erejét gyengíti, hogy nem elképzelhetetlen olyan, az X kutató által alkalmazottnál is szofisztikáltabb csalás, amely akár a hisztogramok, akár a számjegyek eloszlásának vizsgálata során rejtve marad. Jól mutatja ezt, hogy az X által másodikként alkalmazott szofisztikált módszer az egyszerű módszernél mennyivel jobban teljesít az utolsó számjegy teszténél. Ha a feltételezett csaló X kutató jobban ismeri a szóba jöhető csalásellenőrzési módszereket, akkor minden bizonnyal ki is tudja játszani őket. Tud találni olyan megoldást, amelynek segítségével a csalás célja teljesül, de a csalás maga rejtve marad. De ebben az esetben el kell mélyednie a Benford-törvény rejtelseiben, ismernie kell az ezzel foglalkozó kutatási eredményeket is. Minél jobban el kívánja rejteni a csalását, annál inkább. Ez pedig önmagában megdrágítja a csalást, hiszen minél szofisztikáltabb csalást szeretne alkalmazni, hogy a csalása rejtve maradjon, annál több erőfeszítésre lesz szüksége. Következésképpen a Benford-törvény közvetve, ezen az úton is képes csökkenteni a csalás gyakoriságát.

6. ábra: Utolsó számjegy teszt: a számolt χ^2 értékek az eredeti szerződéses értékeknél és a csalást is tartalmazó szerződéses értékeknél, 2005-2023



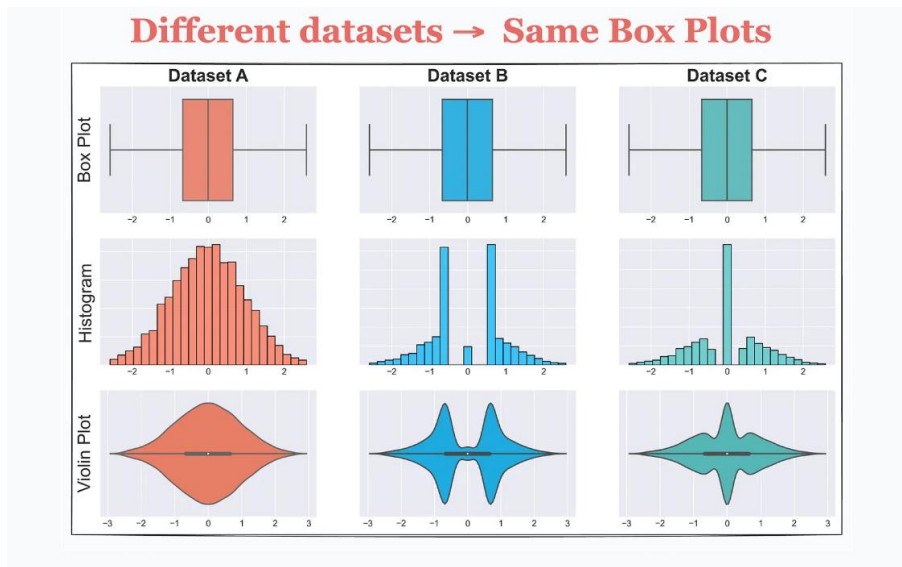
Megjegyzés: keretszerződésekkel együtt; euróban számolt értékek alapján.

Forrás: CRCB

3.17. Ha egy gyanútlan kutató csak az aggregált adatokat és az alapstatisztikákat nézné, akkor szinte semmi különbséget nem látna az eredeti adatok és a két módszerrel meghamisított adatok között. Közismert, hogy önmagában az alapstatisztikákból (átlag, módusz, medián, szórás) nem láthatjuk a vizsgált változó valódi természetét (lásd a 7. ábrát).

Ezt tapasztalhatjuk a bemutatott példánknál is. A top szereplőkhöz, köztük Mészároshoz tartozó cégek által nyert közbeszerzések a hamisított adatokon számolt részesedései sem évente, sem az időszak egészében nem térnek el a tényleges adatokon számoltaktól (lásd a 8. táblázatot).

7. ábra: Többféle adatbázis, azonos box-plot



Forrás: <https://blog.dailydoseofds.com/p/use-box-plots-with-caution-they-can>

8. táblázat: A top szereplőkhöz és közöttük Mészároshoz tartozó cégek által nyert szerződések részesedése az összes szerződésből, 2011-2023, %

	Eredeti adatok		Hamisított adatok, egyszerű eljárás		Hamisított adatok, szofisztikált eljárás	
	Mészáros	Top szereplők	Mészáros	Top szereplők	Mészáros	Top szereplők
2011	0,1	4,9	0,1	4,9	0,1	4,9
2012	0,1	6,2	0,1	6,2	0,1	6,2
2013	0,2	12,8	0,2	12,8	0,2	12,8
2014	1,0	8,0	1,0	8,0	1,0	8,0
2015	0,3	2,1	0,3	2,1	0,3	2,1
2016	3,4	16,4	3,4	16,4	3,4	16,4
2017	7,2	25,8	7,2	25,8	7,2	25,8
2018	4,6	16,2	4,6	16,2	4,6	16,2
2019	7,9	20,9	7,9	20,9	7,9	20,9
2020	9,9	25,0	9,9	25,0	9,9	25,0
2021	15,3	22,5	15,3	22,5	15,3	22,5
2022	12,1	19,8	12,1	19,8	12,1	19,8
2023	3,7	8,5	3,7	8,5	3,7	8,5
Összesen	6,2	15,5	6,2	15,5	6,2	15,5

Megjegyzés: keretszerződésekkel együtt

Forrás: CRCB

3.18. Az X kutató azért csalt, hogy csökkentse az Orbán rendszer top szereplői és az egyéb magyar cégek által nyert közbeszerzések nagyság szerinti megoszlása között ténylegesen fennálló nagymértékű eltérést. Ezért úgy manipulálta az szerződések

nettó értékeinek adatait, hogy egy kiválasztott értékhatár alatt az általa megadott feltételek esetén a szerződések értékét ezen értékhatár fölé tornászta fel, így ezen értékhatár alatt csökkentette, felette pedig növelte a megnyert szerződések arányát az egyéb magyar cégeknél. Miközben, láthattuk, a csalás 2,2 százalékponttal csökkentette az egyéb magyar cégek esetében a kiválasztott értékhatár alatti szerződések arányát (lásd az 5. táblázatot), a leíró statisztikákból nem látszik semmi csalásra utaló jel. Az eredeti és a hamisított adatokra vonatkozó leíró statisztikák a harmadik tizedesjegyig teljesen megegyeznek (lásd a 9. táblázatot).

9. táblázat: A nettó szerződéses érték leíró statisztikái a top szereplők és az egyéb nyertes cégek által nyert szerződéseknél 2015-2023 között az eredeti és a hamisított adatok esetében

	Eredeti adatok	Hamisított adatok, egyszerű eljárás	Hamisított adatok, szofisztikált eljárás
	(v)	(v'_1)	(v'_2)
Top szereplők			
Median	0,446	0,446	0,446
Átlag	6,411	6,411	6,411
SE	0,580	0,580	0,580
N	2835	2835	2835
Egyéb cégek			
Median	0,053	0,053	0,053
Átlag	0,543	0,543	0,543
SE	0,015	0,015	0,015
N	164451	164451	164451
Összes szerződés			
Median	0,055	0,055	0,055
Átlag	0,642	0,642	0,642
SE	0,018	0,018	0,018
N	167286	167286	167286

Megjegyzés: keretszerződésekkel együtt; millió euró.

Forrás: CRCB

4. Összefoglalás és következtetések

4.1. A tanulmány egy hipotetikus csalási példán elemezte az utolsó számjegy teszt – egy, a csalásfelderítési vizsgálatokban alkalmazott eljárás – alkalmazhatóságát és korlátait.

4.2. Először röviden ismertettük Tátrai és Gábos (2025a; 2025b, valamint Tátrai et al., 2025) eredményeit a KSH jövedelemfelvételének 2018-2023 közötti adatainál feltárt anomáliáról, majd a KSH-nak Tátrai és Gábos (2025a; 2025b) elemzésére adott reakcióját. Foglalkoztunk a KSH lehetséges magatartását leíró lehetőségekkel.

4.3. Röviden ismertettük a Világbank Doing Business rangsorának esetét, amelynek tapasztalatait érdemes figyelembe venni akkor, amikor egy adatszolgáltató intézmény adatait megalapozott kritika éri.

4.4. Az elemzés második részében a csalásfelderítés során alkalmazott, a Benford-törvényre alapozódó módszerek alkalmazásának példáiból ismertettünk néhányat. Amellett érveltünk, hogy egy statisztikai jellemző adatsorát vizsgálva érdemes ezeket a módszereket is alkalmazni, mert ezek segíthetnek annak megválaszolásában, hogy létezik-e anomália az adatokban, illetve abban is, hogy a minta milyen egyéb jellemzőihez köthetők ezek az anomáliák. Ezzel pedig az anomáliák okainak felderítéséhez is közelebb juthatunk. Foglalkoztunk e módszerek korlátaival.

4.5. Az elemzésben felvetettük, hogy Tátrai és Gábos (2025a; 2025b) által felfedezett, a KSH jövedelemfelvételére vonatkozó anomáliák vizsgálatánál is hasznos lehet a Benford-törvény adta vizsgálati módszer. Itt mindenképpen figyelembe kell azonban venni, hogy a szóban forgó adatsorok (jövedelemfelvételek) terjedelme elég kicsi, mivel csak a 10 , 10^2 , 10^3 , és az időszak vége felé a 10^4 tartományt, azaz mindössze három, majd négy nagyságrendet fognak át. Feltehető ezért, hogy ezeknél a számjegyek eloszlása nem, illetve csak megközelítőleg lesz a Benford-törvénynek megfelelő. Ezt a problémát azonban kezelni lehet Roukema (2013) által alkalmazott módszerhez analóg módon. Meg lehet figyelni ugyanis például az EU-SILC kutatásban az EU országok 2005-2023 közötti adatsorait és a 2005-2017 közötti – anomáliát nem tartalmazó – magyar adatsorokat és ezekre (több száz mintára) mind kiszámolni azt, hogy a számjegyek empirikus eloszlása mennyiben tér el az elméleti, a Benford-törvény által meghatározott értéktől. Ezzel megadható az is, hogy a kapott χ^2 értékek 95 százalékos valószínűséggel milyen határon belül mozognak. Ha a 2018-2023 közötti magyar adatokra vonatkozó χ^2 értékek ezen a határon túl lennének, akkor ez nyilvánvalóan arra mutatna, hogy a szóban forgó magyar adatok az EU-SILC többi adatsorához képest jelenleg még meg nem határozott okból bekövetkező irregularitást tartalmaznak.

4.6. Az elemzés harmadik részében egy hipotetikus csalási példán mutattuk be az *utolsó számjegy teszt* alkalmazását. A példában egy feltételezett X kutató egy ténylegesen létező sokaság (közbeszerzési szerződések) egy jellemzőjét (a szerződés nettó értéke) manipulálta úgy, hogy egy értékhatár alatt a tényleges adatokat egy,

vagy több értékhatár feletti adatra cserélte ki abból a célból, hogy csökkentse az egyéb magyar cégek súlyát az értékhatár alatti szerződések almintáján. Ezt a csalást a feltételezett X kutató igyekezett a lehető legjobban elrejteni: egyrészt viszonylag kevés szerződés (az összes szerződés 3,2 százaléka) adatait hamisította csak, másrészt olyan szofisztikált technikát is alkalmazott, amellyel, ismerve a Benford-törvényt, még rejtettebbé kívánta tenni az általa elkövetett csalást.

4.7. Mivel a hipotetikus példában szereplő esetben ismerjük a szerződéses érték eredeti (hamis adatokat nem tartalmazó) adatait is, könnyen ábrázolható az, hogy valójában mit is tett a hipotetikus csalás a szerződések eloszlásával. Az utolsó számjegy teszt alkalmazásával pedig számszerűen látszik, hogy – bár az eredeti adatsor egész időszakra számolt értéke számottevően különbözött a Benford-törvény által előírt eloszlástól – a sokaság mindössze 3 százalékát érintő csalás jelentős mértékű újabb irregularitást eredményezett.

4.8. Láthattuk azt is, hogy miközben a csalás csak az esetek töredékét (3,2 százalékát) érintette, valamennyire teljesítette a csaló célját is: csökkentette az „*egyéb magyar cégek*” értékhatár alatti részesedését, így közelítve őket a top szereplőkhöz köthető cégekhez.

4.9. A kapott eredmények alapján az alábbi négy következtetés fogalmazzuk meg:

(1) A Benford-törvényt felhasználó csalásfelderítési módszerek hasznosak lehetnek egy változó elemzésénél, az erre a változóra vonatkozó anomáliák feltárásánál és az anomáliák hatóórainak azonosításánál, bár nem szabad túlbecsülni e módszerek erejét. Egy képzett csaló – ismerve a szóba jöhető csalásfelderítési technikákat – képes lehet elrejteni a csalást. De ebben az esetben ismernie kell a Benford-törvényt és az ezzel foglalkozó kutatási eredményeket is. Minél jobban el kívánja rejteni a csalását, annál inkább. Ez pedig önmagában megdrágítja a csalást, hiszen minél szofisztikáltabb csalást szeretne alkalmazni, hogy a csalása rejtve maradjon, annál több erőfeszítésre lesz szüksége. Következésképpen a Benford-törvény közvetve, ezen az úton is képes csökkenteni a csalás gyakoriságát. Továbbá, figyelembe kell venni, a Benford-törvény által prediktált eloszlástól való eltérés nem jelenti automatikusan azt, hogy csalás következett be. A Benford-törvénytől való eltérésnek a csaláson kívül lehet más oka is.

(2) A bemutatott egészen kis arányú (a minta 2-3 százalékát érintő) csalás is meglehetősen erős hatást gyakorolt az adott változó eloszlására és a változó számjegyeinek eloszlására.

(3) Az értékhatárok adminisztratív kijelölése nemkívánt hatással járhat. A szereplők alkalmazkodnak ezekhez az értékhatárokhoz. Úgy változtatják magatartásukat, hogy a számukra kedvező kimenetet éri el, ami vagy értékhatár alatti, vagy értékhatár feletti adattorlódást fog eredményezni. Az X csaló célja az volt, hogy egy általa megadott értékhatárt figyelembe véve változtassa a változó eredeti eloszlását.

(4) Nyilvánvaló szabály, hogy egy kutatónak, ha az elemi adatok rendelkezésére állnak, a leíró statisztikák kiszámolásán túl ajánlott első lépésként megvizsgálni annak az változónak eloszlását is, amelyet az elemzés során fel kíván használni.

Tátrai és Gábos (2025a; 2025b) által feltárt anomáliák és a KSH-nak ezekre adott reakciója (KSH, 2025) nyomán, úgy tűnik, ennek a követelménynek a KSH által szolgáltatott minden adatállomány és indikátor esetében érdemes eleget tenni.

Hivatkozások

Anonimus. (n.d.). *A Benford-törvény, avagy meghamisították-e az adatainkat?*

<https://people.inf.elte.hu/kiss/statisztika/8ea.pdf>

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572. <https://www.jstor.org/stable/984802>

Benko-Kiss, A. (2020). Néhány adatbázis elemzése a Benford-törvény segítségével.

In Kis, K., Komarek, L., Monostori, T. (Eds) *Mezőgazdasági és vidékfejlesztési kutatások a jövő szolgálatában*. (pp. 39-45). MTA SZAB Mezőgazdasági Szakbizottság.

https://www.researchgate.net/profile/Arpad-Benko-Kiss/publication/349063824_NEHANY_ADATBAZIS_ELEMZESE_A_BENFORD-TORVENY_SEGITSEGEVEL_Benk_Kiss_Arpad/links/601da282a6fdcc37a8062b66/NEHANY-ADATBAZIS-ELEMZESE-A-BENFORD-TOERVENY-SEGITSEGEVEL-Benk-Kiss-Arpad.pdf

https://www.researchgate.net/profile/Arpad-Benko-Kiss/publication/349063824_NEHANY_ADATBAZIS_ELEMZESE_A_BENFORD-TORVENY_SEGITSEGEVEL_Benk_Kiss_Arpad/links/601da282a6fdcc37a8062b66/NEHANY-ADATBAZIS-ELEMZESE-A-BENFORD-TOERVENY-SEGITSEGEVEL-Benk-Kiss-Arpad.pdf

https://www.researchgate.net/profile/Arpad-Benko-Kiss/publication/349063824_NEHANY_ADATBAZIS_ELEMZESE_A_BENFORD-TORVENY_SEGITSEGEVEL_Benk_Kiss_Arpad/links/601da282a6fdcc37a8062b66/NEHANY-ADATBAZIS-ELEMZESE-A-BENFORD-TOERVENY-SEGITSEGEVEL-Benk-Kiss-Arpad.pdf

Charslaw, Ch. (1988). Anomalies in Income Numbers: Evidence of Goal Oriented Behavior. *The Accounting Review*, 63(2), 321-327.

<https://www.jstor.org/stable/248109>

Diekmann, A., & Jann, B. (2010). Benford's Law and Fraud Detection: Facts and Legends. *German Economic Review*, 11(3), 397–401.

<https://doi.org/10.1111/j.1468-0475.2010.00510.x>

Dlugosz, S., & Müller-Funk, U. (2009). The value of the last digit: statistical fraud detection with digit analysis. *Advances in Data Analysis and Classification*, 3, 281–290.

<https://doi.org/10.1007/s11634-009-0048-5>

Eckhardt, G. M., & Ruxton, G. D. (2023). Investigating and preventing scientific misconduct using Benford's Law. *Research Integrity and Peer Review*, 8, 1.

<https://doi.org/10.1186/s41073-022-00126-w>

Élő, A. (2025a, Április 2). Valami nagyon nem stimmel a KSH szegénységi adataival.

Válasz Online. <https://www.valaszonline.hu/2025/04/02/ksh-eurostat-szegenyseg-inflacio-jovedelmek-megbizhatosag/>

Élő, A. (2025b, Április 4). KSH-botrány: a tudósok szerint gyenge lábakon áll a hivatal védekezése.

Válasz Online. <https://www.valaszonline.hu/2025/04/04/ksh-szegenysegi-adatok-botrany-valasz/>

Fassbender, C., Houde, S., Silver-Balbus, S., Ballard, K., Kim, B., Rutledge, K. J., Dixon, J. F., Iosif, A. M., Schweitzer, J. B., & McClure, S.M. (2014). The decimal effect: behavioral and neural bases for a novel influence on intertemporal choice in healthy individuals and in ADHD. *Journal of Cognitive Neuroscience*, 26(11), 2455–2468.

https://doi.org/10.1162/jocn_a_00642

R. M. Fewster, R. M. (2009). A Simple Explanation of Benford's Law. *The American Statistician*, 63(1), 26-32.

Gideon, M., Helppie-McFall, B., & Hsu, J. W. (2017). *Heaping at Round Numbers on Financial Questions: The Role of Satisficing*. The Federal Reserve.

<https://www.federalreserve.gov/econresdata/feds/2017/files/2017006pap.pdf>

Gyürky, Gy. & Farkas, J. (2015). Az első számjegyek Benford-törvénye és a rádióaktív izotópok felezési ideje. *Fizikai Szemle*, 65(9), 297-301. https://real-j.mtak.hu/3749/38/FizSzem-2015_09.pdf

Honda, H., Kagawa, R., & Shirasuna, M. (2022). On the round number bias and wisdom of crowds in different response formats for numerical estimation. *Sci Rep* 12(8167). <https://doi.org/10.1038/s41598-022-11900-7>

Isaac, M. S., Wang, Y., & Schindler, R. M. (2021). The Round-Number Advantage in Consumer Debt Payoff. *Journal of Cognitive Neuroscience*, 31(2), 240-262.

<https://doi.org/10.1002/jcpy.1192>

Kossofsky, A. E., (2015). *Benford's Law. Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications*. World Scientific.

<https://doi.org/10.1142/9089>

Központi Statisztikai Hivatal. (2025, Április 3). A Központi Statisztikai Hivatal válasza a Valami nagyon nem stimmel a KSH szegénységi adataival című, a Válasz Online felületén 2025. április 2-án megjelent cikkben jelzett állításokra. [Sajtóközlemény].

https://www.ksh.hu/sajtoszoba_kozlemenyek_tajekoztatok_2025_04_03

Machen, R. C., Jones, M. T., Varghese, G. P., & Stark, E. L. (2021). Investigation of Data Irregularities in Doing Business 2018 and Doing Business 2020. Investigation Findings and Report to the Board of Executive Directors (September 15, 2021). WilmerHale.

<https://thedocs.worldbank.org/en/doc/84a922cc9273b7b120d49ad3b9e9d3f9-0090012021/original/DB-Investigation-Findings-and-Report-to-the-Board-of-Executive-Directors-September-15-2021.pdf>

Mebane, W.R. (2010). Fraud in the 2009 presidential election in Iran? *CHANCE* 23, 6-15. <https://doi.org/10.1007/s00144-010-0003-4>

Milicz., Á. (2016). Kvantitatív módszerekkel támogatott üzemgazdasági ellenőrzések, különös tekintettel a Benford-törvény alkalmazhatóságára. *Statisztikai Szemle*, 94(6), 612-634.

Miller, S. J. (2015). A Quick Introduction to Benford's Law. In S. J. Miller (Ed.), *Benford's Law. Theory & Applications* (pp. 3-22). Princeton University Press.

<https://press.princeton.edu/books/hardcover/9780691147611/benfords-law>

Murphy, K. M, Shleifer, A., & Vishny, R. W. (1993). Why Is Rent-Seeking So Costly to Growth? *American Economic Review Papers and Proceedings*, 83(2), 409-414.
<https://www.jstor.org/stable/2117699>

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*. 4(1), 39–40.
<https://doi.org/10.2307/2369148>

Nigrini, M. J., & Mittermaier, L. J. (1997). The use of Benford's Law as an aid in analytical procedures. *Auditing: a journal of practice & theory*, 16(2), 52-67.

Nigrini, M. J. (2012). *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Wiley. <https://www.wiley.com/en-us/Benford's+Law%3A+Applications+for+Forensic+Accounting%2C+Auditing%2C+and+Fraud+Detection-p-9781118152850>

Nigrini, M. J. (2015). Detecting Fraud and Errors Using Benford's Law. In S. J. Miller (Ed.), *Benford's Law. Theory & Applications* (pp. 191-211). Princeton University Press.
<https://press.princeton.edu/books/hardcover/9780691147611/benfords-law>

Pena-Marin, J., & Bhargave, R. (2016). Lasting performance: Round numbers activate associations of stability and increase perceived length of product benefits. *Journal of Consumer Psychology*, 26(3), 315-459.
<https://doi.org/10.1016/j.jcps.2015.11.004>

Rauch, B., Götsche, M., Brähler, G., & Engel, S. (2015). Measuring the Quality of European Statistics. In S. J. Miller (Ed.), *Benford's Law. Theory & Applications* (pp. 235-243). Princeton University Press.
<https://press.princeton.edu/books/hardcover/9780691147611/benfords-law>

Randall Morck, R., & Shou J. Ch. (2018). *On the Integrity of the "Ease of Doing Business" Indicators*. (Final report, June 18, 2018). doingbusiness.org.
https://www.doingbusiness.org/content/dam/doingBusiness/media/Miscellaneous/Audit/Doing-Business-External-Audit-2018_Full-Report.pdf

Reuters. (2021, September 16). Probe found 'undue pressure' from World Bank leaders to boost China business rankings. *Reuters*.
<https://www.reuters.com/world/china/probe-found-undue-pressure-world-bank-leaders-boost-china-business-rankings-2021-09-16/>

Roukema, B. F. (2013). A first-digit anomaly in the 2009 Iranian presidential election. *Journal of Applied Statistics*, 41(1), 164–199.
<https://doi.org/10.1080/02664763.2013.838664>

Roukema, B. F. (2015). Complementing Benford's Law for Small *N*: A Local Bootstrap. In S. J. Miller (Ed.), *Benford's Law. Theory & Applications* (pp. 223-232). Princeton University Press.
<https://press.princeton.edu/books/hardcover/9780691147611/benfords-law>

- Rundlett, A., & Svolik, M. W. (2016). Deliver the Vote! Micromotives and Macrobehavior in Electoral Fraud. *American Political Science Review*, 110(1), 180-197. <https://doi.org/10.1017/S0003055415000635>
- Sandefur, J., & Wadhwa, D. (2018, February 5). A Change in World Bank Methodology (Not Reform) Explains India's Rise in Doing Business Rankings. *Center for Global Development*. <https://www.cgdev.org/blog/change-world-bank-methodology-not-reform-explains-indias-rise-doing-business>
- Skovoroda, R., & Lankina, T. (2016). Fabricating votes for Putin: new tests of fraud and electoral manipulations from Russia. *Post-Soviet Affairs*, 33(2), 100-123. <https://doi.org/10.1080/1060586X.2016.1207988>
- Tátrai, A., & Gábos, A. (2022). A jövedelmi szegénység kiterjedtsége és mélysége 2005 és 2020 között – eltérő küszöbértékek mentén. In Kolosi, T., Szelényi, I. & Tóth I. Gy. (Eds.), *Társadalmi Riport 2022* (pp. 243-260). Társi Zrt. https://tarki.hu/sites/default/files/2022-12/243_260TRIP2022_GabosTatrai.pdf
- Tátrai, A. (2025, October 5-9). *The reliability of EU-SILC income data in Hungary* [Conference presentation abstract]. 65th ISI World Statistics Congress, The Hague, the Netherlands. <https://www.isi-next.org/abstracts/submission/2534/view/>
- Tátrai, A., & Gábos, A. (2025a, Április 1). Kutatói aggodalmak az EU-SILC magyarországi jövedelemadataival kapcsolatban. [Researchers' concerns about EU-SILC income data for Hungary]. *Válasz Online*. <https://www.valaszonline.hu/2025/04/01/eu-silc-magyarorszag-i-jovedelemadok-kutato-i-elemzes/>
- Tátrai, A., & Gábos, A. (2025b, Április 11). Researcher Concerns About the Reliability of EU-SILC Income Data in Hungary. *Válasz Online*. <https://www.valaszonline.hu/2025/04/11/reliability-eu-silc-income-data-in-hungary/>
- Tátrai, A., Gábos, A., Huszár, Á., Krekó, J. & Tomka, Zs. (2025). Data Quality Issues in Hungarian Income Data in EU-SILC, 2014-2023 V3. https://www.researchgate.net/profile/Annamaria-Tatrai-2/publication/392362865_Data_Quality_Issues_in_Hungarian_Income_Data_in_EU-SILC_2014-2023_V3/links/683eed04c33afe388ac9d836/Data-Quality-Issues-in-Hungarian-Income-Data-in-EU-SILC-2014-2023-V3.pdf
- The Editorial Board. (2021, September 30). The China World Bank Scandal. Beijing allegedly intervened to improve its business ranking. *The Wall Street Journal*. <https://www.wsj.com/opinion/the-china-world-bank-scandal-doing-business-kristalina-georgieva-11633039063>
- Tóth, I. J., & Hajdu, M. (2017). *Versenyerősség, társadalmi veszteségek és a 25 millió forintos értékhatár rejtélye. 2009-2016 közötti magyar közbeszerzések*

statisztikai vizsgálata. CRCB https://www.crcb.eu/wp-content/uploads/2017/09/t25_2017_paper_170927.pdf

Tóth, G. (2021). A számviteli beszámolók minőségének fejlesztési lehetőségei a könyvvizsgálat szabályozásán keresztül. [PhD értekezés, Magyar Agrár és Élettudományi Egyetem]. DOI: 10.54598/000610. https://phd.mater.uni-mate.hu/57/2/toth_gabor_tezis_DOI.pdf

Tóth, I. J., & Hajdu, M. (2022). Cronyism in Orbán Regime: An Empirical Analysis of Public Tenders, 2005-2021. In Csanádi, M., Gerő, M., Hajdu, M., Kovách, I., Laki, M., & Tóth, I. J., *Dynamics of an Authoritarian System. Hungary, 2010-2021* (pp. 230-274). Central European University Press.

Tödter, K-H. (2009). Benford's Law as an Indicator of Fraud in Economics. *German Economic Review*, 10(3), 339–351. <https://doi.org/10.1111/j.1468-0475.2009.00475.x>

Tödter, K-H. (2015). Benford's Law and Fraud in Economic Research. In S. J. Miller (Ed.), *Benford's Law. Theory & Applications* (pp. 244-256). Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691147611/benfords-law>

Varian, H. (1972). Benford's Law. *The American Statistician*, 26(3), 65-66. <https://www.jstor.org/stable/2682871>

Wadhwa M., & Zhang K. (2015). This Number Just Feels Right: The Impact of Roundedness of Price Numbers on Product Evaluations. *Journal of Consumer Research*, 41(5), 1172-1185. <https://doi.org/10.1086/678484>

Wadhwa, M., & Zhang, K. (2019). When numbers make you feel: Impact of round versus precise numbers on preventive health behaviors. *Organizational Behavior and Human Decision Processes*, 150, 101-111. <https://doi.org/10.1016/j.obhdp.2018.08.005>.

Wang, H., Lu, T., Zhang, Y., Wu, Y., Sum, Y., Dong, J., & Huang, W. (2025). Last digit tendency: Lucky numbers and psychological rounding in mobile transactions. *Fundamental Research*, 5(1), 370-378. <https://doi.org/10.1016/j.fmre.2023.11.011>.

World Bank Group (2018a). *World Bank Group Statement on External Audit of Chile's Rankings in Doing Business report* (Statement 2018, July 18). The World Bank. <https://www.worldbank.org/en/news/statement/2018/07/11/world-bank-group-statement-on-external-audit-of-chiles-rankings-in-doing-business-report>

World Bank Group (2018b). *Report on the Integrity of the "Ease of Doing Business" Indicators: Management Response*. (2018, July 18). World Bank Group. https://www.doingbusiness.org/content/dam/doingBusiness/media/Miscellaneous/Audit/Doing-Business-External-Audit-2018_Action-Plan.pdf

World Bank Group (2020a). Doing Business Archive. About Us. *World Bank Group*.
<https://archive.doingbusiness.org/en/about-us>

World Bank Group (2020b). About Doing Business. *World Bank Group*.
<https://openknowledge.worldbank.org/server/api/core/bitstreams/0be02f45-ade4-551d-a85d-986be83e77f0/content>

World Bank Group (2020c). Doing Business – Data Irregularities Statement (Statement, August 27, 2020). *World Bank Group*.
<https://www.worldbank.org/en/news/statement/2020/08/27/doing-business---data-irregularities-statement>

World Bank Group (2021a). World Bank Group to Discontinue Doing Business Report (Statement, September 16, 2021). *World Bank Group*.
<https://www.worldbank.org/en/news/statement/2021/09/16/world-bank-group-to-discontinue-doing-business-report>

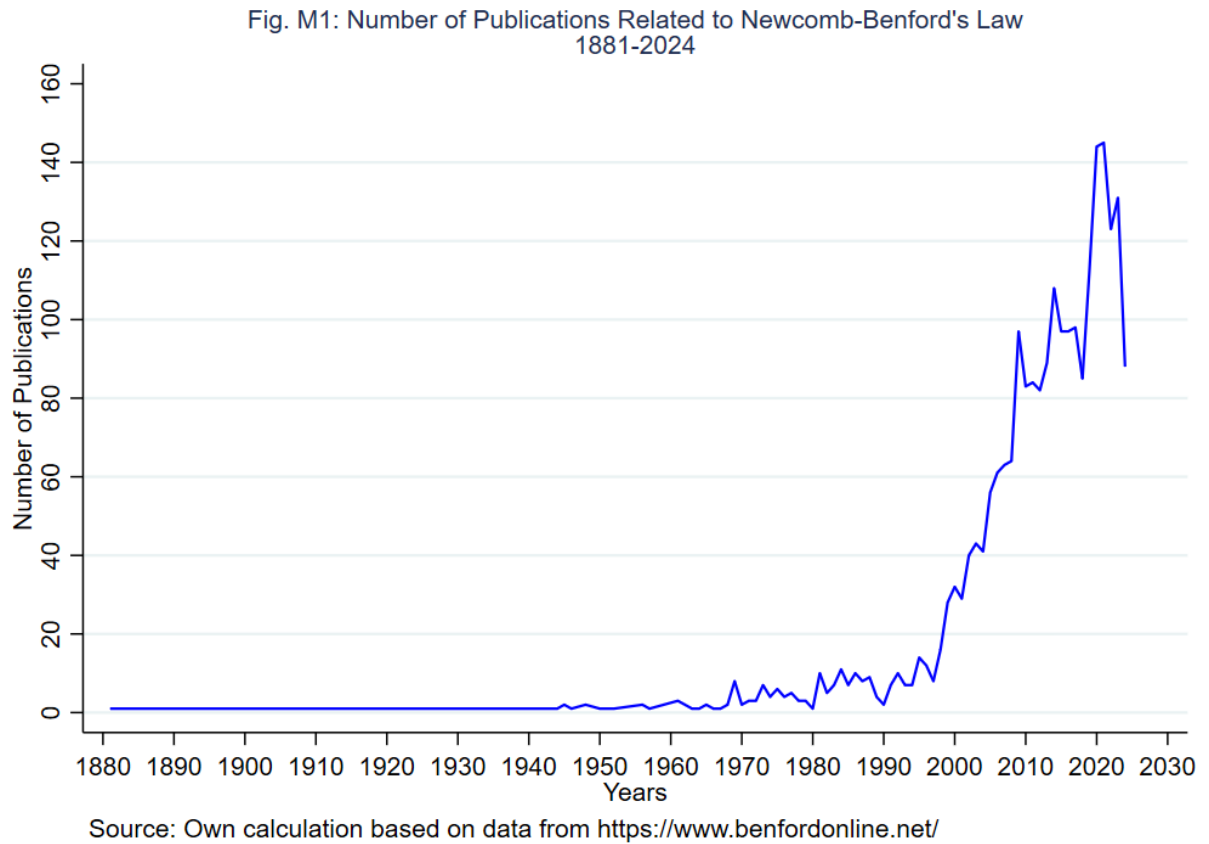
World Bank Group (2021b). Doing Business Archive. *World Bank Group*.
<https://archive.doingbusiness.org/en/doingbusiness>

Zumbrun, J., & Talley, I. (2018, January 12). World Bank Unfairly Influenced Its Own Competitiveness Rankings The World Bank's chief economist said he would recalculate national rankings of business competitiveness going back at least four years. *The Wall Street Journal*. <https://www.wsj.com/articles/world-bank-unfairly-influenced-its-own-competitiveness-rankings-1515797620>

Zumbrun, J. (2020, August 27). World Bank Halts Report on National Competitiveness Rankings Amid Concerns of Data Manipulation. China, Azerbaijan, UAE and Saudi Arabia among nations whose data may have been inappropriately altered. *The Wall Street Journal*. <https://www.wsj.com/world/world-bank-delays-report-on-national-competitiveness-rankings-amid-concerns-of-data-manipulation-11598554654>

Melléklet

M1. A Newcomb-Benford-törvénnyel foglalkozó közlemények száma 1881-2024



M2. A Newcomb-Benford-törvénnyel foglalkozó közlemények listája 2025. május 20-ig a <https://www.benfordonline.net> oldalon:

<https://www.crcb.eu/?p=3669>

M3. A Világbank Doing Business rangsorának esete

A Doing Business rangsora az országok jogrendszerének intézményeinek versenyképességét mérte. Azt, hogy az adott ország jogrendje mennyire segíti elő, könnyíti meg a vállalkozások indítását, piacon való működését és a piacról való kilépését. A rangsor összehasonlítható adatokat közölt mintegy 190 országra vonatkozóan. (World Bank Group, 2020a; World Bank Group, 2020b). A rangsor készítését 2002-ben indították el és 2003-ban hozták nyilvánosságra az első jelentést. A rangsorral kapcsolatos anomáliákra elemzők először 2018 januárban hívták fel a figyelmet (Zumbrun & Talley, 2018). Ezt követte egy külső szakértő vizsgálat, amely cáfolta az anomáliákra vonatkozó feltételezéseket (Randall & Shou, 2018). A szakértői jelentés nyilvánosságra hozása után a Világbank sajtóközleményt bocsátott ki, amely üdvözölte a vizsgálat eredményeit, ezek figyelembe vételét ígérte a későbbiekben, valamint idézte az anomáliák kiküszöbölésére szolgáló akcióterv lépéseit is (World Bank Group, 2018a; World Bank Group, 2018b):

"We welcome the thoughtful and thorough findings of the Doing Business external review, and we are gratified to see that the review found the previous allegations concerning the integrity of the report and the broader Doing Business team to be unfounded and without merit.

We are closely considering the recommendations from the reviewers and will take steps to implement improvements as necessary. A management response on the recommendations is available [here](#).

The Doing Business Index remains a unique and invaluable tool for countries looking to improve their business climate, and we believe that going forward it will be an even stronger and more vital report."

[World Bank Group (2018a)]

A szakmai közvéleményt azonban nem nyugtatták meg sem az első vizsgálat eredményei, sem a Világbank állásfoglalása, sem az általa közzétett akcióterv. Ezek után került sor a 2019-es rangsor közzétételére 2020-ban. Szakmai kritikák újra anomáliákra és feltételezhető adatmanipulációkra hívták fel a figyelmet (Zumbrun, 2020). Feltételezések szerint Kína, Azerbajdzsán és az Egyesült Arab Emírségek adatait manipulálhatták, "tették kedvezőbbé" a rangsor készítői. A kritikák hatására a Világbank kénytelen volt a Doing Business rangsor felfüggesztése mellett dönteni (World Bank Group, 2020c) a tervek szerint addig, amíg egy mélyebb szakmai vizsgálatot ("systematic review and assessment of data changes that occurred subsequent to the institutional data review process for the last five Doing Business reports") le nem folytatnak:

"Over the 17 years of its existence, the Doing Business report has been a valued tool for countries seeking to measure costs of doing business. Doing

Business indicators and methodology are designed with no single country in mind, but rather to help to improve the overall business climate.

A number of irregularities have been reported regarding changes to the data in the Doing Business 2018 and Doing Business 2020 reports, published in October 2017 and 2019. The changes in the data were inconsistent with the Doing Business methodology.

The integrity and impartiality of our data and analysis is paramount and so we are immediately taking the following actions:

- We are conducting a systematic review and assessment of data changes that occurred subsequent to the institutional data review process for the last five Doing Business reports.*
- We have asked the World Bank Group's independent Internal Audit function to perform an audit of the processes for data collection and review for Doing Business and the controls to safeguard data integrity.*

We will act based on the findings and will retrospectively correct the data of countries that were most affected by the irregularities.

The Board of Executive Directors of the World Bank has been briefed on the situation as have the authorities of the countries that were most affected by the data irregularities.

The publication of the Doing Business report will be paused as we conduct our assessment."

[World Bank Group (2020c)]

A független vizsgálat 2021 nyarán zárult le. Ennek eredményeit közlő jelentést (Machen et al., 2021) a Világbank nyilvánosságra hozta. A jelentés a rangsor előállításáért felelős vezetők beavatkozását, etikátlan magatartását állapította meg. A jelentés szerint a rangsort készítő csoportra a Világban egyes vezetői nyomást gyakoroltak azért, hogy a rangsort Kína számára előnyösebbé tegyék, meghamisítsák:

"The changes to China's data in Doing Business 2018 appear to be the product of two distinct type of pressure applied by Bank leadership ont he Doing Business team: (1) pressure—both direct and indirect— applied by senior staff in the Office of the President presumably at the direction of President Kim to change the report's methodology in an effort to boost China's score: and (2) pressure applied by CEO Georgieva and her advisor, Mr. Djankov, to make specific change to China's data points in an effort to increase its ranking at precisely the same time the country was expected to play a key role in Bank's capital increase campaign."

[Machen et al. (2021)]

A jelentés főbb megállapításait a nemzetközi sajtó (például az Economist, The Wall Street Journal, Financial Times, Reuters) is ismertette:

"An investigation conducted by law firm Wilmerhale concluded that leaders of the World Bank applied "undue pressure" to secure changes aimed at improving China's ranking in the bank's "Doing Business 2018" and those of other countries in the 2020 report.

It said it found that then-CEO Kristalina Georgieva, now the managing director of the International Monetary Fund, and a key adviser pressured staff to "make specific changes to China's data points" and boost its ranking, at a time when the bank was seeking China's support for a big capital increase.

Georgieva, in a statement, rejected the findings of the probe and said she had met with the IMF's executive board to discuss the matter."

[Reuters (2021)]

A Világbank miután visszamenőleg revideálta a korábban általa közölt adatokat és a revideált adatokat nyilvánosságra hozta, úgy döntött, hogy véglegesen lezárja a Doing Business projektet, nem publikál új Doing Business rangsort (World Bank Group, 2021a; 2021b):

"Trust in the research of the World Bank Group is vital. World Bank Group research informs the actions of policymakers, helps countries make better-informed decisions, and allows stakeholders to measure economic and social improvements more accurately. Such research has also been a valuable tool for the private sector, civil society, academia, journalists, and others, broadening understanding of global issues.

After data irregularities on Doing Business 2018 and 2020 were reported internally in June 2020, World Bank management paused the next Doing Business report and initiated a series of reviews and audits of the report and its methodology. In addition, because the internal reports raised ethical matters, including the conduct of former Board officials as well as current and/or former Bank staff, management reported the allegations to the Bank's appropriate internal accountability mechanisms.

After reviewing all the information available to date on Doing Business, including the findings of past reviews, audits, and the report the Bank released today on behalf of the Board of Executive Directors, World Bank Group management has taken the decision to discontinue the Doing Business report. The World Bank Group remains firmly committed to advancing the role of the private sector in development and providing support to governments to design the regulatory environment that supports this. Going forward, we will be working on a new approach to assessing the business and investment climate. We are deeply grateful to the efforts of the many staff members who

have worked diligently to advance the business climate agenda, and we look forward to harnessing their energies and abilities in new ways.”

[World Bank Group (2021a)]

A Világbank azzal, hogy saját honlapján nyilvánosságra hozta a Doing Business kutatásra vonatkozó független jelentést, fontos lépést tett hitelességének védelmében. Talán, ha a rangsort ért első kritikákra nem sietve és egy viszonylag felületes vizsgálattal válaszolt volna, meg tudta volna akadályozni a későbbi adatmanipulációs kísérleteket is és új szabályokkal ellent tudott volna állni a politikai nyomásgyakorlásnak is. Ekkor talán a Doing Business rangsor is megmenekülhetett volna.