



UNIWERSYTET
WARSZAWSKI

WYDZIAŁ
SOCJOLOGII

Jakub Rybacki

Wydział Socjologii Uniwersytetu Warszawskiego, Katedra Socjologii Cyfrowej

MET 2023, Warszawa 05.07.2023



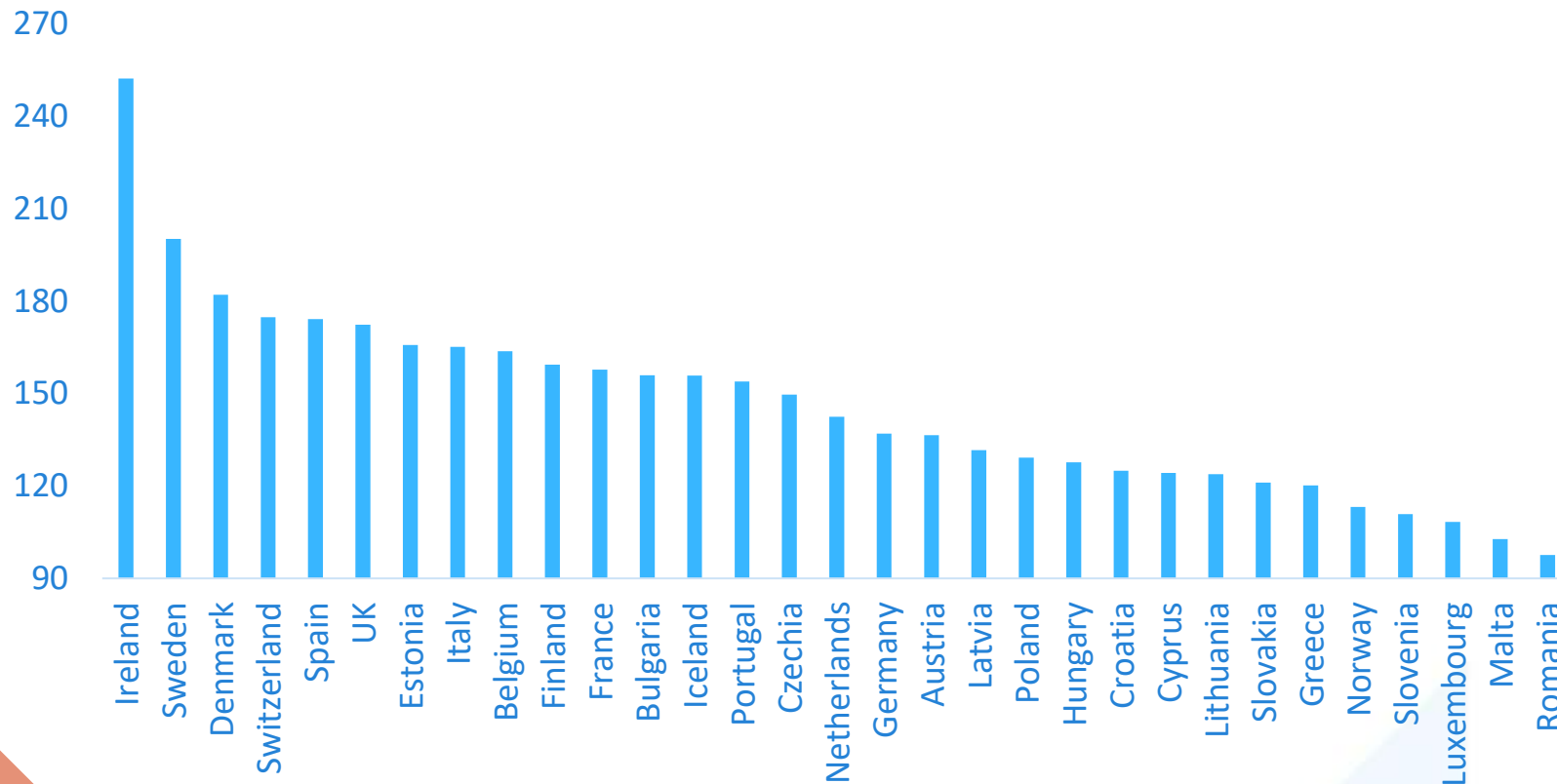
Podsumowanie badania:

1. Główny wniosek: Statystyka publiczna przedstawia błędny obraz zmian cen elektroniki RTV.
2. Dane: Oferty cen promocyjnych z lat 2017 – 2023 pobrane z portalu pepper.pl
3. Metoda: Regresja hedoniczna uwzględniająca podstawowe cechy produktów.



Sprzęt ICT - Brak harmonizacji danych

HICP index - Information processing equipment in January 2010



Źródło: [Eurostat](#)

Sprzęt ICT - Brak harmonizacji danych

Table: HICP subindices - Audio-visual, photographic and information processing equipment

Category	Audio-visual, photographic and information processing equipment	Equipment for the reception, recording and reproduction of sound and picture	Photographic and cinematographic equipment and optical instruments	Information processing equipment
COICOP	CP0911	CP0912	CP0913	CP0914
Amplitude	78.9	133.3	192.6	154.5
Min	97.9	99.9	95.3	97.7
Max	176.8	233.2	287.9	252.2

Source: Own calculations based on the Eurostat data



Literatura

1. Rozbieżności między krajami potrafią być duże:

Broda and Weinstein (2007): Wzrost CPI Japonia będzie o 0.6 pkt proc. mniejszy rocznie, jeżeli zastosujemy metodologię z USA do przeliczeń

2. Techniki hedoniczne zalecane przy szybko zmieniających się produktach:

[Haan and Diewert \(2017\)](#): Różnice rzędu 40 pkt proc. w 18 lat w zależności od wykorzystanej metody.

3. Wykorzystanie do elektroniki rozwijane przez urzędy statystyczne np. przez [INSEE](#) (2019)



Prosty model hedoniczny

Formuła tzw. **Time Dummy Hedonic Indexes** (np., Diewert et al. 2007):

$$\ln(p_{n,t}) = A_t + \sum K_{n,t} + e_{n,t}$$

Gdzie:

- A_t - wektor obrazujący średni poziom cen w kolejnych okresach
- $K_{n,t}$ - macierz cech produktów, mierzone są różnice względem rynkowego standardu
- $e_{n,t}$ - błąd losowy



Alternatywne źródło danych – Pepper

1. Portal z ogłoszeniami dotyczącymi promocji – dane pozyskane przez webscrapping
2. Archiwalne oferty z dużą częstotliwością od 2018 r.







Table: Pepper.pl – number of observation for specific components

Component.	Graphic Cards	SSD drives	RAM memory cards
2016	49	127	24
2017	95	94	57
2018	231	575	126
2019	444	563	290
2020	478	714	252
2022	2171	691	276
2023 (till mid-June)	500	336	107



Czemu to może zadziałać?

Canon EOS 2000D czarny + 18-55mm IS II - Pozostałe oferty

 128 opinii	CANON EOS 2000D + EF-S 18-55 mm f/3.5-5.6 IS II	1 838,00 zł DARMOWA WYSYŁKA dostępny	IDŹ DO SKLEPU
 272 opinii	Aparat Canon 2000D + 18-55 IS II Canon Polska 24 miesiące gwarancji GWARANCJA 2 LATA / FAKTURA VAT	1 839,00 zł Z wysyłką od 1861,00 zł dostępny	IDŹ DO SKLEPU
 223 opinii	APARAT CANON EOS 2000D + 18-55 IS II	1 859,00 zł DARMOWA WYSYŁKA dostępny	IDŹ DO SKLEPU
 1867 opinii	Canon EOS 2000D + EF-S 18-55mm f/3,5-5.6 IS II - RABAT NA STRONIE - Kup na Raty - RRSO 0% DARMOWY TRANSPORT OD 99 ZŁ!	1 889,00 zł DARMOWA WYSYŁKA dostępny	IDŹ DO SKLEPU
 639 opinii	Canon EOS 2000D + EF-S 18-55mm f/3,5-5.6 IS II - szybka wysyłka! DARMOWY TRANSPORT OD 99 ZŁ	1 889,00 zł DARMOWA WYSYŁKA dostępny	IDŹ DO SKLEPU
 798 opinii	Aparat CANON EOS 2000D BK 18-55 mm IS II EU26 DARMOWY TRANSPORT ZAMÓW DZIŚ, ODBIERZ JUTRO	1 946,91 zł DARMOWA WYSYŁKA dostępny	IDŹ DO SKLEPU

Źródło: [Ceneo.pl](https://ceneo.pl)



Oczyszczanie danych.

1. Usuwanie produktów błędnie dopasowanych do kategorii.
2. Portale uzupełniają dane za użytkowników na podstawie treści ich ogłoszeń – rezultaty czasem są ciekawe.
3. Przykład: Kość RAM vs. Kość wołowa.



Źródło: [Wikipedia](#)



Źródło: [PixaBay](#)





Identyfikacja cech produktów.

- *Name Entity Detection* – proste wyrażenia regularne:
 - Stosunkowo prosto określić producenta oraz model urządzenia (95%+ przypadków)
 - Tylko część ofert będzie miała przedstawione parametry sprzętu - 60-80% ogłoszeń w zależności od kategorii.
 - Badane komponenty zawierają pojedyncze parametry – w przypadku np. laptopów będzie potrzebne bardziej zaawansowane narzędzie.
- Imputacje:
 - Algorytm K-najbliższych sąsiadów – porównanie do najbliższego sąsiada.
 - Todo: Sprawdzić skuteczność dopasowania na próbie walidacyjnej.



Karty graficzne - przykład

Table: Price evolution of exemplary Radeon GPU card

Year	2018	2019	2020	2021	2022	2023
Radeon RX 5800	1,221	729	768	2,058	710	
Radeon RX 6800			3,743	5,098	3,357	2,599
Average price	1,395	1,357	1,638	3,244	3,042	2,941

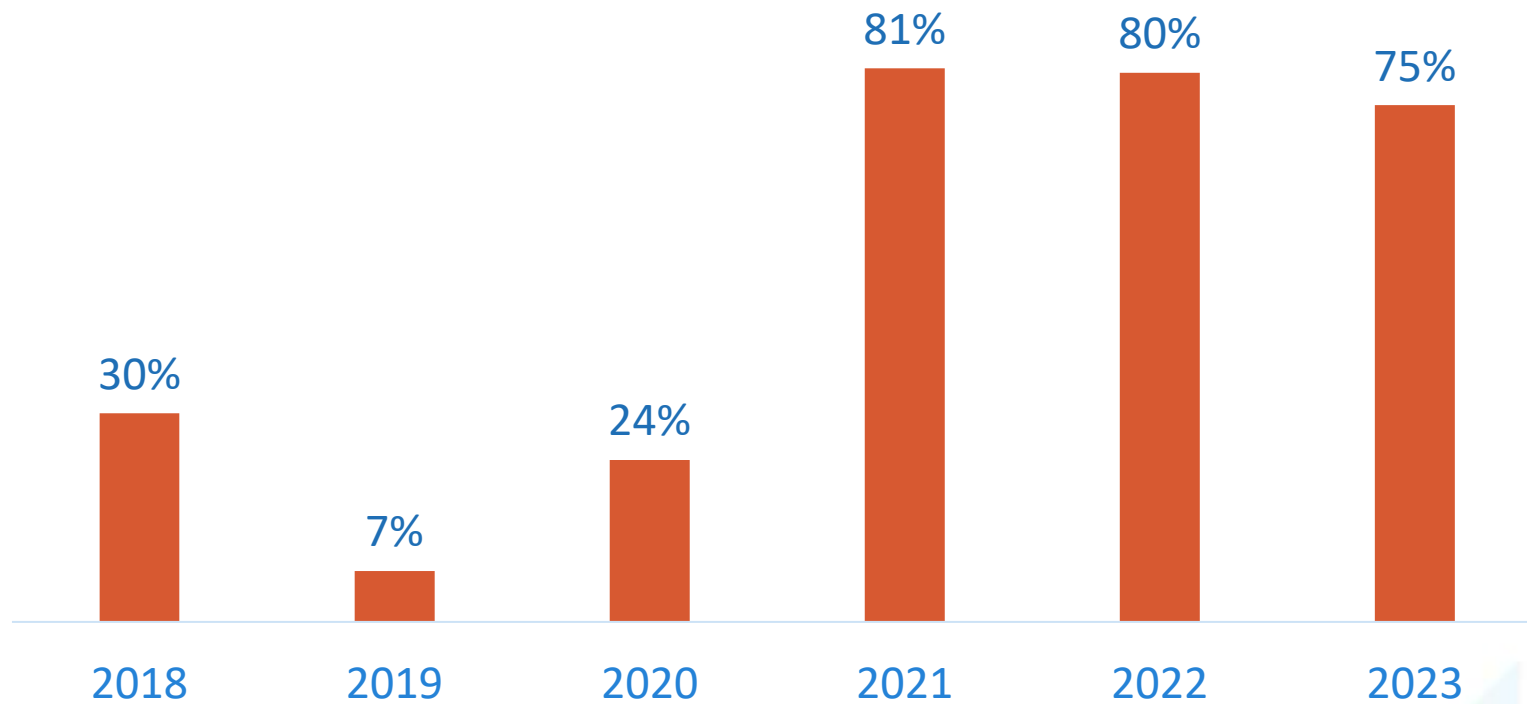
Karty graficzne - regresja

Variable	Coefficient	Standard error	T statistic	P> t
Model constant	6.73	0.07	96.29	0.00
Days from first offer	0.00	0.00	-9.08	0.00
Quality Difference in GB	0.06	0.00	28.90	0.00
.....				
2018	0.30	0.05	5.64	0.00
2019	0.07	0.05	1.48	0.14
2020	0.24	0.05	4.67	0.00
2021	0.81	0.05	16.06	0.00
2022	0.80	0.05	16.66	0.00
2023	0.75	0.05	14.46	0.00



Karty GPU – ceny wzrosły od 2017 r.

GPU prices increase since 2017



Dysk SSD - przykład

Table: Price evolution of exemplary SSD hard drive – Crucial MX 500

Size (GB) / Year	2018	2019	2020	2021	2022	2023
250	204.0	473.0	176.6	151.3	159.6	133.8
500	321.9	262.3	253.7	225.9	216.2	184.0
1024	651.8	470.1	432.4	393.3	363.1	299.9
2048	1,181.4	850.3	779.2	796.5	729.1	540.4
Average Price	488.0	405.4	390.5	378.6	439.6	410.1

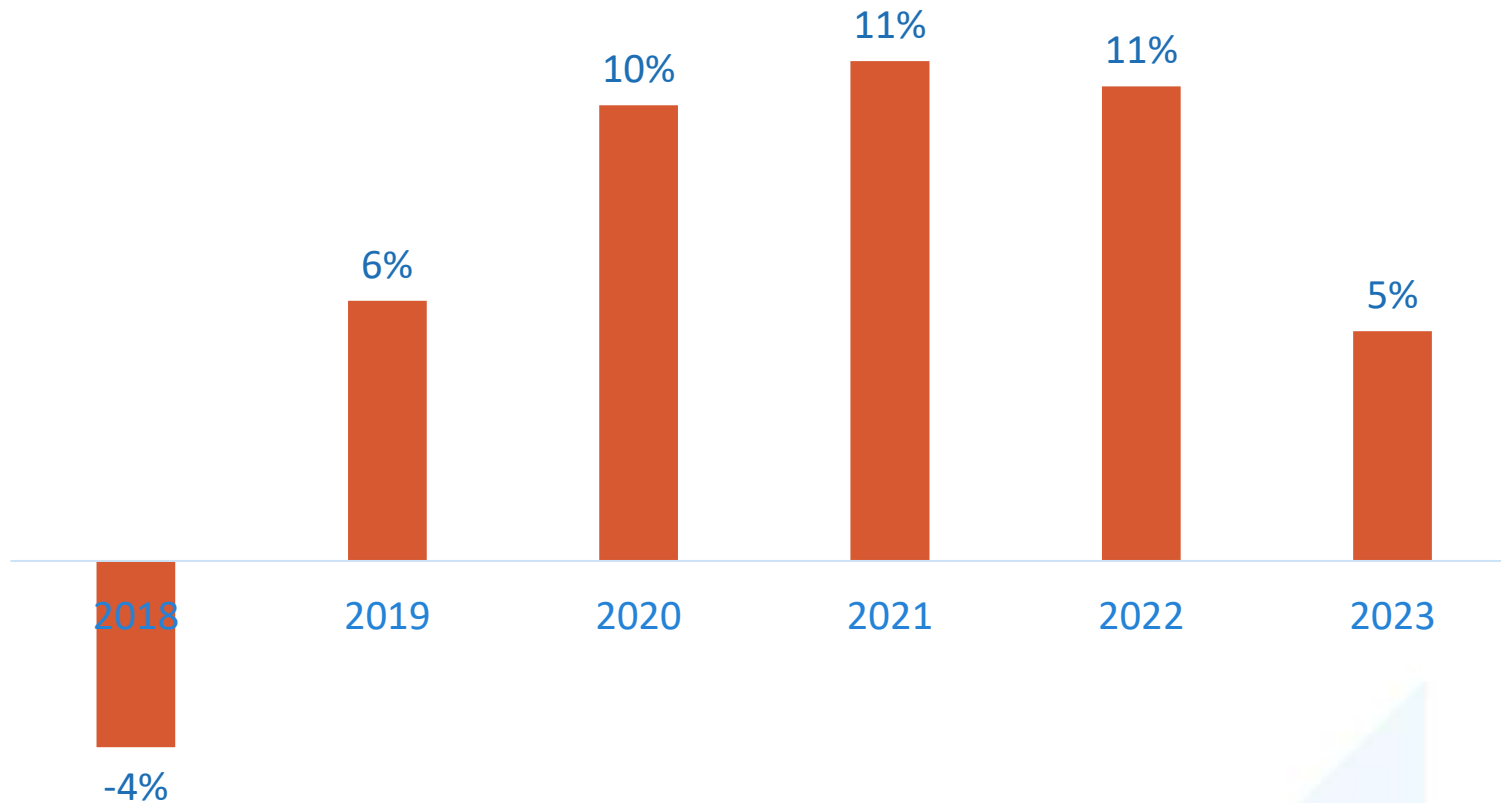
Dysk SSD - Regresja

	Coefficient	Standard error	T statistic	P> t
Model constant	5.91	0.07	87.15	0.00
Quality Difference (no of GB)	0.00	0.00	13.40	0.00
.....				
2018	-0.04	0.06	-0.69	0.49
2019	0.06	0.06	0.96	0.34
2020	0.10	0.06	1.75	0.08
2021	0.11	0.06	1.91	0.06
2022	0.11	0.06	1.80	0.07
2023	0.05	0.06	0.80	0.43

Uwaga: Regresja wymaga dopracowania - różnice cenowe warunkowane są np. przez typ złącza M.2 czy SATA.

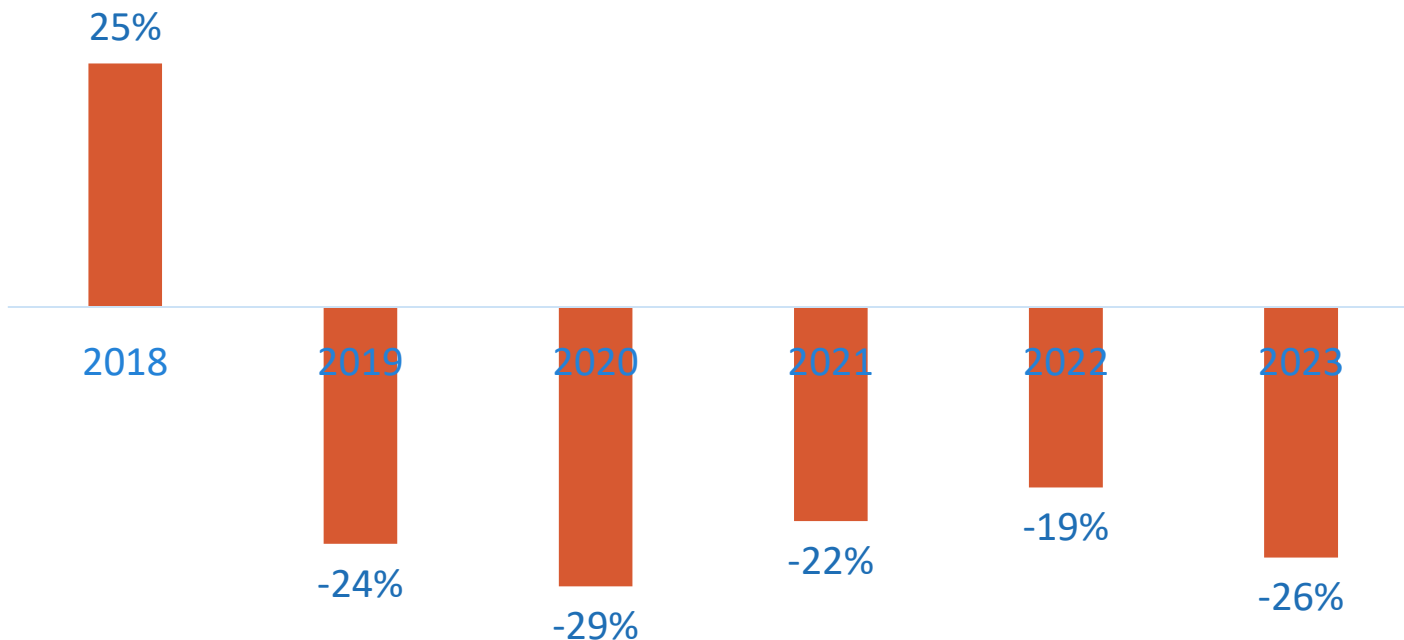
Dyski SSD – ceny wzrosły od 2017 r.

SSD prices increase since 2017



Kości RAM – ceny spadły od 2017 r.

RAM prices increase since 2017



Uwaga: Regresja może wymagać dopracowania



Podsumowanie

1. Regresja, która uwzględnia rozwój technologiczny i charakterystyki producentów prowadzi do zupełnie odmiennych wniosków niż widoczne w statystyce publicznej.
2. Ceny elektroniki systematycznie rosną
3. Potencjalne elementy do dopracowania
 - Zmniejszyć odchylenia standardowe estymowanych parametrów
 - Dopracowanie rozpoznawania cech produktów
 - Analityka błędów imputacji

