

SEMINARIUM

Pomiar inflacji

Stan obecny i wyzwania



METODOLOGIA POMIARU INFLACJI: STAN OBECNY I WYZWANIA

Jacek Biątek
Uniwersytet Łódzki
Główny Urząd Statystyczny

Warszawa, grudzień 2019

Agenda:

- Rys historyczny pomiaru inflacji – Polska i świat
- Metodologia badania cen konsumpcyjnych
- Źródła danych o cenach
- Źródła danych o wydatkach konsumentów
- Formuły indeksowe w pomiarze inflacji
- Procedura (etapy) liczenia inflacji
- Źródła obciążenia pomiaru wskaźnika cen towarów i usług konsumpcyjnych
- Zakres zastosowań wskaźnika inflacji
- Alternatywne źródła danych w pomiarze inflacji
- Wyzwania towarzyszące nowym źródłom danych
- Przykładowe analizy wykorzystująca nowe źródła danych

Rys historyczny pomiaru inflacji – Polska i świat

Pierwsza próba stworzenia **wskaźnika cen** sięga początków XVIII w. W **1707** roku brytyjski biskup **Wiliam Fleetwood** opublikował w swoim dziele *Chronicon Preciosum* prosty koszyk dóbr, dzięki któremu porównywał poziom cen z 1460 i 1707 roku. Porównywanymi dobrami były **4 beczki piwa, 5 kwart zboża i 6 jardów tkaniny**.

Pierwszy **nieważony indeks** w historii został zaproponowany (formalnie) przez francuskiego ekonomistę **Nicolasa Dutot (1738)**. Dutot przyjął stosunek średnich cen stałego koszyka w okresie t do okresu bazowego, aby sprawdzić realną deflację przychodów króla Ludwika XII (dane z 1515 roku) oraz Ludwika XV (dane z 1735 roku). W koszyku Dutot zamieścił sporą liczbę dóbr (m.in. **belka siana, gołąb, koza, królik, kura**), **ale również co ciekawe usługi (dzienna pensja kobiet oraz mężczyzn)**.

Włoski ekonomista i astronom **Giovanni Carli** wpadł na nieco inny pomysł ćwierć wieku później. W **1764** roku na podstawie obserwacji cen **zboża, wina i oliwy** zaproponował formułę średniej arytmetycznej stosunku cen relatywnych z okresu t do okresu bazowego. (2.1)

Najbardziej popularna formuła elementarna to jednak indeks **Jevonsa**. W **1863** roku brytyjski ekonomista **William Stanley Jevons** zaproponował rozwiązanie podobne do indeksu Dutot, jednak zamiast średniej arytmetycznej zasugerował przyjęcie średniej geometrycznej z indeksów cząstkowych. Z aksjomatycznego i ekonomicznego punktu widzenia jest to najlepsza formuła elementarna.

Następnie pojawia się idea łączenia dóbr i usług w podgrupy i grupy, które mają przecięż różne udziały na rynku konsumenckim. Pojawia się idea indeksów agregatowych, ich aksjomatyki (**Fisher, 1922**). Oprócz mnogości formuł indeksowych (o tym później) powstają **teoretyczne podstawy dla COLI i CPI**.

Wskaźniki zmian cen (indeksy cen) towarów i usług konsumpcyjnych są obliczane przez Główny Urząd Statystyczny od **1950** roku (na najniższym poziomie – formuła **Jevonsa**, na najwyższych – formuła **Laspeyresa**). Od tego roku są też dostępne dane o rocznych wskaźnikach zmian cen konsumpcyjnych. Od **1982** roku rozpoczęto obliczanie wskaźników miesięcznych.

Metodologia badania cen konsumpcyjnych w Polsce

Wskaźniki cen detalicznych (CPI, HICP)

CPI

Indeksem, który odzwierciedla zmiany kosztów konsumpcji gospodarstw domowych przy jednoczesnym utrzymaniu użyteczności koszyka na stałym poziomie jest **indeks kosztów utrzymania** (COLI – Cost of Living Index, indeks Könusa).

W praktyce każdy konsument ma indywidualną **funkcję użyteczności**, która w dodatku nie jest bezpośrednio mierzalna. Za najlepsze przybliżenie COLI uchodzą tzw. **indeksy superlatywne**, niemniej jednak w praktyce – z uwagi na opóźnienie publikacji wag koszyka – indeks kosztów utrzymania **przybliża się indeksem CPI opartym na formule Laspeyresa**. W tym miejscu należy jednak zaznaczyć, iż **koszyk CPI jest nieco węższy** niż teoretyczny koszyk COLI (np. wskaźnik COLI uwzględnia również czynniki środowiskowe wpływające na ogólny poziom życia, jak jakość wody czy koszty walki z przestępczością).
Inne kraje liczą np. CPI-U, C-CPI, RPI, RPIJ, etc.

HICP

Ponieważ lokalnie funkcjonujące indeksy CPI (różniące się metodologicznie pomiędzy krajami jak i różniące pod względem zawartości koszyka dóbr) są nieodpowiednie do międzynarodowych porównań, z inicjatywy Europejskiego Banku Centralnego (**EBC**) powołano do życia indeks **HICP (Harmonised Index of Consumer Prices)**, dedykowanego pomiarowi inflacji w krajach strefy Euro.

Każdy kraj UE musi stosować tę samą metodologię narzuconą przez standardy HICP co gwarantuje wspomnianą porównywalność między krajami. Wartości wskazań indeksów HICP uzyskane od państw członkowskich stanowią również bazę dla globalnego indeksu dynamiki cen w całej UE, czyli indeksu **MUICP (Monetary Union Index of Consumer Prices)**.

Formuła: łańcuchowa wersja **indeksu typu Laspeyresa z grudniem każdego roku jako „miesiącem-łącznikiem”**.

Koszyk towarów i usług konsumpcyjnych oraz jego klasyfikacja

Punktem wyjścia do ustalenia zawartości „koszyka” reprezentantów towarów i usług konsumpcyjnych jest klasyfikacja spożycia indywidualnego według celu (COICOP/HICP) opracowana przez **ONZ**.

Do **2013** r. była stosowana przez GUS **4-cyfrowa klasyfikacja COICOP**. Od 2013 r. w GUS rozpoczęto proces rozszerzania klasyfikacji COICOP, wprowadzając 5-ty, wspólny dla krajów UE, szczebel klasyfikacji (podklasy) w związku z wdrażaniem tej klasyfikacji COICOP przez Eurostat. Jednocześnie w ramach modyfikacji klasyfikacji COICOP uwzględniono **nowe towary i usługi**, które pojawiły się na rynku w ciągu kilkunastu lat stosowania klasyfikacji COICOP (np. w zakresie sprzętu teleinformatycznego).

Nowa, rozszerzona europejska klasyfikacja składa się z **dwunastu** 2-cyfrowych głównych działów (ECOICOP 2):

- 01 – Żywność i napoje bezalkoholowe
- 02 – Napoje alkoholowe i wyroby tytoniowe
- 03 – Odzież i obuwie
- 04 – Użytkowanie mieszkania lub domu i nośniki energii
- 05 – Wyposażenie mieszkania i prowadzenie gospodarstwa domowego
- 06 – Zdrowie
- 07 – Transport
- 08 – Łączność
- 09 – Rekreacja i kultura
- 10 – Edukacja
- 11 – Restauracje i hotele
- 12 – Pozostałe wydatki na towary i usługi.

Kolejnymi szczeblami klasyfikacji ECOICOP są:

- 3-cyfrowy – 44 grup,
- 4-cyfrowy – 110 klas,
- 5-cyfrowy – 296 podklas.

W ramach niektórych podklas GUS utworzył dodatkowy 6-cyfrowy poziom krajowy ze względu na wymogi metodyczne w zakresie obliczenia wskaźników cen oraz potrzeby użytkowników. Podklasy oraz klasy, które nie posiadają podklas, tworzą około **340 grup elementarnych** obejmujących jednorodne towary i usługi, dla których dostępne są wagi z badania budżetów gospodarstw domowych.

Wybór reprezentantów do koszyka inflacyjnego

Podstawowymi kryteriami doboru reprezentantów grup elementarnych są **dominujący udział wartości sprzedaży** towaru lub usługi w łącznej wartości sprzedaży grupy elementarnej, którą mają reprezentować, oraz **powszechność ich sprzedaży wśród konsumentów**.

Przy doborze reprezentantów wykorzystywane są:

- obserwacje ankietatorów notujących ich ceny;
- analizy i ekspertyzy prowadzone przez GUS
- zapisy respondentów uczestniczących w badaniach budżetów gospodarstw domowych;
- analizy danych uzyskiwanych od ich gestorów.

Aktualnie obserwacji podlegają ceny detaliczne około **1500** reprezentantów towarów i usług konsumpcyjnych. Lista towarów i usług reprezentantów jest stała przez cały rok i obowiązująca dla notowań cen w całym kraju.

W ramach listy reprezentantów możemy wyróżnić ich dwa rodzaje z punktu widzenia charakteru ich opisu, a mianowicie:

- **reprezentanty opisane w sposób precyzyjny**, z podaniem konkretnych parametrów ściśle określających ich cechy (nazwa, typ, parametry techniczno-eksploatacyjne, jednostka miary);
- **reprezentanty opisane w sposób ogólny**, stanowiące wąskie grupy asortymentowe. Reprezentują one np. ubiory, bieliznę, obuwie itp., które są dostarczane na rynek w krótkich seriach. O wyborze konkretnych reprezentantów takich grup towarów lub usług decyduje ankieter.

Lista towarów i usług reprezentantów podlega **corocznej aktualizacji**, stosownie do zmieniającego się modelu konsumpcji oraz zmian na rynku detalicznym.

Modyfikację listy wymuszają także **zmiany klasyfikacyjne** oraz konieczność dostosowania badań do **standardów międzynarodowych**. Może dojść do rozszerzenia listy bądź też zastąpieniu dotychczasowego reprezentanta nowym.

Zasady uwzględniania produktów sezonowych

Za **produkty sezonowe** uznawane są te produkty, które nie są dostępne dla konsumentów lub sprzedawana ich ilość jest zbyt mała w pewnych cyklicznie powtarzających się okresach każdego roku (np. niektóre **warzywa i owoce** lub **odzież letnia czy zimowa**). Wyróżniamy **słabą i silną sezonowość**.

Dla produktu sezonowego w **pierwszym miesiącu** poza sezonem przyjmowana jest jego tzw. **typowa cena (bazowa)** obliczana jako średnia arytmetyczna z cen tego produktu zanotowanych w sezonie (np. w zakresie odzieży i obuwia z wyłączeniem pierwszego i ostatniego miesiąca sezonu).

W kolejnych miesiącach poza sezonem (począwszy od drugiego miesiąca) jako wskaźniki cen produktów sezonowych przyjmowane są **średnie z miesięcznych wskaźników cen** produktów reprezentantów, zaliczanych do tych samych grup elementarnych lub agregatowych ECOICOP, występujących w tym okresie lub dostępnych na rynku przez cały rok.

Zasady uwzględniania zmian jakości produktów

W sytuacji, gdy dotychczasowy produkt (towar lub usługa) reprezentant został wycofany ze sprzedaży lub trudno byłoby utrzymać ciągłość notowań jego cen następuje zmiana tego reprezentanta.

Sytuacja 1. Nowy produkt reprezentant może być uznany za **równoważny** dla dotychczasowego reprezentanta, gdy spełnione są następujące warunki:

- jest wyrobem tego samego producenta,
- ma ten sam skład surowcowy oraz parametry techniczne i jakościowe, istotne dla nabywcy,
- ma tę samą jednostkę sprzedaży lub różnica w tym zakresie jest nieistotna,
- ma ten sam rodzaj opakowania,
- różnice są nieistotne dla nabywcy i dotyczą tylko cech związanych z gustem i upodobaniami konsumenta.

Następuje tu bezpośrednie porównanie cen (*direct price comparison*).

Sytuacja 2. Stwierdzamy, iż nowy i stary produkt reprezentant **nie są sobie równoważne** (nastąpiła **zmiana jakości produktu**).

Typowe błędy i ich skutki:

- **niewwzględnianie** zmian jakościowych prowadzi do przeszacowania inflacji, jeżeli ma miejsce poprawa jakości (HICP manual: *Implicit Quality Index IQI*);
- traktowanie zmian wynikających z podążania za **modą** jako zmian jakości prowadzi do niedoszacowania inflacji.

W celu „wyceny” zmiany jakości (**quality adjustment**) stosuje się metody bezpośrednio **explicit methods** (tu wykorzystuje się **charakterystyki produktu**) lub pośrednio **implicit methods** (wykorzystuje się różnice lub ilorazy cen produktów przyjętych za **ekwiwalentne** i zakłada, że różnice te wynikają z różnic jakościowych). Aby adekwatnie ustalić nową ceną stosowany jest **współczynnik korygujący**.

Przykładowe metody **bezpośrednie**: (a) wycena zmiany wielkości opakowania (*package size adjustment*); (b) wycena dodatkowych opcji (*option pricing*); (c) regresja hedonistyczna (*hedonic regression*).

Przykładowe metody **pośrednie**: (a) zakładka pomostowa (*bridged overlap*); (b) imputacja cen bazowych (*backcasting*)

Zasady uwzględniania obniżek cen i rabatów

W przypadku, gdy towar jest sprzedawany po **obniżonej cenie** jego cena jest notowana jedynie wtedy gdy dotyczy towaru **pełnowartościowego**. Obniżona cena jest notowana bez względu na długość jej stosowania, tj. nawet wtedy gdy obniżona cena obowiązuje 1 dzień i właśnie w tym dniu, zgodnie z harmonogramem, przypada jej notowanie.

Jeżeli przy zakupie stosowany jest **rabat** jest on uwzględniany w notowaniach cen tylko wtedy gdy jest stosowany **powszechnie**, tj. oferowany jest każdemu potencjalnemu nabywcy bez żadnych specjalnych warunków, niezależnie od wielkości dokonywanych zakupów i może być przypisany do zakupu pojedynczego produktu.

Źródła danych o cenach

Notowania cen dokonywane przez ankieterów „w terenie”

Notowania cen są prowadzone przez ankieterów Urzędów Statystycznych w około **35 tysiącach punktów notowań** cen zlokalizowanych w **207 rejonach** badania na terenie całego kraju.

Rejonami są miasta lub części dużego miasta (np. gminy lub dzielnice), na obszarze których znajdują się punkty notowań. Badania cen konsumpcyjnych na terenach wiejskich nie są prowadzone ze względu na wysokie koszty oraz niewielki asortyment towarów i usług, który jest tam dostępny. Doboru rejonów dokonuje w sposób **celowy** GUS przy **współdziale urzędów statystycznych**. Liczba rejonów podlega corocznej weryfikacji. Punktami notowań cen są punkty sprzedaży detalicznej, placówki gastronomiczne lub jednostki świadczące usługi: **sklepy powszechne, sklepy wyspecjalizowane, domy towarowe i handlowe, supermarkety i hipermarkety, apteki, stacje paliw oraz stałe punkty sprzedaży drobnodetalicznej (kioski, stragany), bary, restauracje, kawiarnie, piwiarnie, punkty gastronomiczne, stołówki etc.**

Centralne notowania cen

Ceny detaliczne reprezentantów towarów i usług notowane centralnie przez pracowników Urzędów Statystycznych dotyczą tych reprezentantów, które mają **cenę jednolite obowiązującą na terenie całego kraju lub jego części**, ustalane przez organy administracji rządowej i organy jednostek samorządu terytorialnego oraz podmioty prowadzące działalność gospodarczą. Notowania tych cen prowadzone są na podstawie **cenników**, zarządzeń i decyzji tych organów oraz podmiotów prowadzących działalność gospodarczą. Ceny notowane centralnie dotyczą reprezentantów towarów i usług m.in. w następujących grupach elementarnych:

- **Energia elektryczna**
- **Wyroby farmaceutyczne**
- **Transport pasażerski kolejowy, morski i śródlądowy**
- **Loty krajowe i międzynarodowe**
- **Usługi telefonii stacjonarnej i komórkowej oraz usługi internetowe**
- **Gry hazardowe**
- **Gazety i czasopisma**

Notowania cen towarów i usług kupowanych w Internecie

Obok notowań cen towarów reprezentantów dokonywanych w punktach notowań przez ankieterów prowadzone są także równoległe dla szeregu grup elementarnych, notowania cen **reprezentantów** zakupywanych przez Internet. Przykładowe grupy elementarne objęte tym badaniem to:

ryż, mąka i pozostałe zboża, mleko pełne świeże, mleko świeże niskotłuszczowe, jogurt, cukier, czekolada, kawa, herbata, samochody osobowe używane, turystyka zorganizowana za granicą, urządzenia elektryczne do higieny osobistej, urządzenia nonelektryczne do higieny osobistej, artykuły do higieny osobistej i kosmetyki, zegary i zegarki, artykuły dla zwierząt domowych, komputery osobiste, wyroby farmaceutyczne, odzież męska, odzież damska, odzież dziecięca, obuwie męskie, obuwie damskie, obuwie dziecięce, itd.

Gestorzy danych

GUS pozyskuje dane wykorzystywane do tworzenia próby i obliczeń indeksów cen detalicznych z systemów informatycznych **sieci handlowych (w tym stacje paliw)** oraz **zakładów ubezpieczeń**.

- Informacje przekazywane przez **sieci handlowe** są aktualnie wykorzystywane **wyłącznie do weryfikacji reprezentatywności produktów** reprezentantów dla grup elementarnych i ewentualnej podmiany produktów reprezentantów na inne oraz wprowadzania do notowań cen nowych produktów reprezentantów.
- Dane przekazywane przez **stacje paliw** dotyczą następujących grup elementarnych: olej napędowy, benzyna, gaz ciekły, smary, oleje, płyny. (Planowane włączenie do CPI w 2020 r.)
- Natomiast dane przekazywane przez **zakłady ubezpieczeń** odnoszą się do poniższych grup: ubezpieczenia mieszkaniowe, prywatne ubezpieczenia zdrowotne, ubezpieczenia pojazdów silnikowych, pozostałe ubezpieczenia. (Do 2019 r. w obliczeniach CPI wykorzystywane dane w zakresie ubezpieczeń pojazdów.)

Źródła danych o wydatkach konsumentów

Badania budżetów gospodarstw domowych

Badania budżetów gospodarstw domowych prowadzone są **metodą reprezentacyjną** co umożliwia uogólnianie uzyskanych wyników z określoną dokładnością na wszystkie prywatne gospodarstwa domowe w kraju.

W badaniu stosowana jest **metoda rotacji całkowitej**, z miesięcznym okresem wymiany gospodarstw domowych, czyli w każdym miesiącu w badaniu uczestniczą inne gospodarstwa domowe.

W każdym miesiącu badaniem objętych jest ponad **3,1 tys.** mieszkań, czyli w każdym roku ponad **37 tys.** mieszkań wylosowanych do próby, a w wylosowanych mieszkaniach badane są **wszystkie gospodarstwa domowe**.

Stosowana jest **50-procentowa rotacja gospodarstw w następnym roku**, tzn. 50% tych samych gospodarstw domowych uczestniczy w badaniu w danym miesiącu w kolejnych dwóch latach.

Głównym źródłem danych o rozchodach każdego badanego gospodarstwa domowego, które to dane są wykorzystywane do opracowania systemu wag, jest „**książeczka budżetowa**” prowadzona przez gospodarstwo domowe w formie papierowej lub elektronicznej przez miesiąc oraz papierowe paragony dostarczane ankietarowi przez gospodarstwa domowe wypełniające książeczkę. Książeczkę budżetową w formie papierowej wypełnia około **40%** gospodarstw domowych, w formie elektronicznej około **2%**, a około **58%** gospodarstw przekazuje informacje o swoich wydatkach na konkretne towary i usługi konsumpcyjne w formie paragonów.

Dodatkowym źródłem jest kwestionariusz „**Karta statystyczna gospodarstwa domowego**” (wydatki użytkownika mieszkania/domu).

Firmy badające rynek

Alternatywa: **Nielsen** lub **GfK**



Formuły indeksowe w pomiarze inflacji

Indeksy bilateralne

- **Formuły nieważone:** Dutot, Carli, Jevons, CSWD, BMW;
- **Formuły ważne:** Laspeyres, Paasche, Fisher, Törnqvist, Walsh, Marshall-Edgeworth, Vartia, Sato-Vartia, Young, Lowe, Palgrave, Davies, Stuvell, Banajree, AG, Lloyd-Moutlon, Divisia;
- **Indeksy łańcuchowe.**

Indeksy multilateralne: GEKS, CCDI, JGEKS, Geary-Khamis, TPD, FBEW + metody aktualizacji oknem czasowym (*window splice, movement splice, half splice, mean splice*).

Nurty w teorii indeksów cen: podejście aksjomatyczne, ekonomiczne, stochastyczne (tzw. stare i nowe), czynnikowe, addytywne, łańcuchowe.

Procedura liczenia inflacji

1) ustalenie przeciętnych cen produktu reprezentanta w rejonie:

a) gdy cenę notuje ankieter dwukrotnie (warzywa, owoce) obliczana jest średnia arytmetyczna z notowań; generalnie jednak cena reprezentanta jest notowana raz i traktuje się jako przeciętną miesięczną cenę w rejonie;

b) gdy cenę notuje ankieter ale dodatkowo pobiera się ją również z Internetu, stosowana jest średnia arytmetyczna ważona z tych cen (wagi pochodzą z BD);

c) w przypadku cen jednolitych oraz taryf, przeciętne miesięczne ceny detaliczne produktów reprezentantów obliczane są jako średnie arytmetyczne ważone liczbą dni obowiązywania poszczególnych cen

d) w przypadku gestorów danych (*zakłady ubezpieczeń*) zamiast przeciętnych cen w rejonie oblicza się indeksy cen na poziomie krajowym;

2) **indeks cen reprezentanta w rejonie** obliczany jako stosunek jego średniej ceny w danym miesiącu w rejonie do jego średniej ceny z okresu bazowego.

3) **ogólnopolskie indeksy cen** produktów reprezentantów obliczane są jako średnie geometryczne wskaźników cen produktów reprezentantów ze wszystkich rejonów (formuła Jevonsa);

4) **indeksy cen** detalicznych towarów i usług konsumpcyjnych **na wyższych szczeblach agregacji** obliczane są za pomocą formuły **Laspeyresa**.

Źródła obciążenia pomiaru CPI

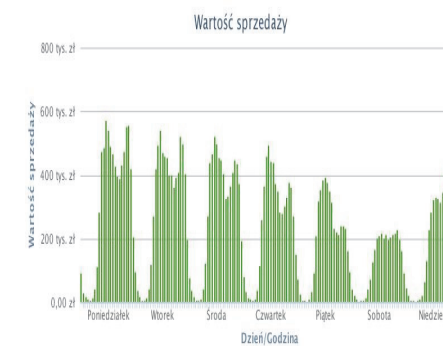
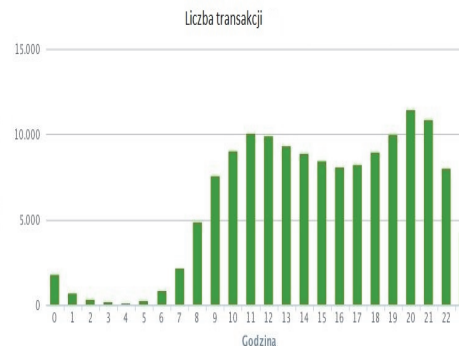
- (1) **Obciążenie wynikające z substytucji dóbr** (*commodity substitution bias*). Obciążenie to wynika ze zmian relatywnych cen poszczególnych dóbr wchodzących w skład koszyka CPI. Efekt substytucji polega na tym, że konsumenci reagują na zmiany cen poprzez zamianę tych dóbr lub usług konsumpcyjnych, które są relatywnie droższe, na dobra relatywnie tańsze.
- (2) **Obciążenie wynikające z substytucji rynku zbytu** (*outlet substitution bias*). Obciążenie to wynika z migracji konsumentów w kierunku atrakcyjniejszym, często właśnie się pojawiającym, rynkom dla zakupów. Takim nowym rynkiem może być np. hurtownia internetowa czy punkt sprzedaży wysyłkowej.
- (3) **Obciążenie wynikające z pojawiania się nowych dóbr** (*new goods bias*). Źródłem tego rodzaju obciążenia są nowe dobra, z jakich w okresie objętym badaniem inflacji, zaczęli korzystać konsumenci. Najczęściej są to produkty dotąd innowacyjne, powstałe na skutek wprowadzenia nowej technologii ich wyrobu, a które weszły właśnie do powszechnego użycia.
- (4) **Obciążenie wynikające ze zmian jakości produktów** (*quality adjustment bias*). Jest to ten rodzaj obciążenia szacunków CPI, który wynika ze zmieniającej się (wraz np. z rosnącymi oczekiwaniami klientów-konsumentów) jakości oferowanych przez rynek dóbr.
- (5) **Obciążenie wynikające z przyjętej formuły kalkulacji** (*formula bias*).

Zakres zastosowań wskaźnika inflacji

- Pomiar inflacji (tempo zmian siły nabywczej pieniądza w czasie);
- Strategia bezpośredniego celu inflacyjnego NBP;
- Waloryzacja wynagrodzeń, emerytur, transferów socjalnych;
- Deflator dla szeregów czasowych;
- Indeksacja umów (finansowych, prawnych);
- Cele analityczne (NPV, PI, kredyty).

Alternatywne źródła danych w pomiarze inflacji

- Dane skanowane
- Dane skrapowane



The screenshot shows the OBI website interface for 'Tynki' (plaster). The browser's developer tools are open, displaying the HTML structure of the product listing. Key elements highlighted in the code include:

- A red box around a product description: `<p>Baumaster Tynk mozaikowy kolor BM545</p>`
- A green box around a rating element: ` (3)`
- An orange box around a price element: ` == $0 `
- A blue box around a unit price element: ``

- Dane transakcyjne z elektronicznych platform handlowych

Wyzwania towarzyszące nowym źródłom danych

- umowy z sieciami;
- dobór próby;
- klasyfikacja produktów do grup COICOP;
- dopasowanie produktów (*matching*);
- filtrowanie produktów;
- imputacja cen (ilości);
- budowa IT;
- wybór formuły indeksu;
- system wag?

Przykładowa analiza wykorzystująca nowe źródła danych

- analiza **danych skanowanych** (dane miesięczne z sieci: XXX, okres grudzień 2017 – sierpień 2019, 212 punktów sprzedaży);
- grupa mleko (180730 rekordów, 142 zidentyfikowane produkty o różnym kodzie producenta **prodID**):
 - mleko zagęszczone i w proszku (19 produktów, 52552 rekordów);
 - mleko kozie (4 produkty, 4463 rekordów);
 - mleko pełne świeże pasteryzowane (17 produktów, 27781 rekordów);
 - mleko pełne świeże UHT (28 produktów, 27948 rekordów);
 - mleko niskotłuszczowe (54 produkty, 34264 rekordów);
 - mleko niskotłuszczowe UHT (20 rekordów, 33722 rekordów)

Dodatkowa analiza:

- „przeżywalność” punktów sprzedaży na całym odcinku czasu:

0.9903846

- „przeżywalność” kodów producenta na całym odcinku czasu:

0.7207207;

- „przeżywalność” kodów producenta w grupach:

mleko kozie (0.5), mleko zagęszczone i w proszku (0.82), mleko pełne świeże pasteryzowane (0.58), mleko pełne świeże UHT (0.8), mleko niskotłuszczowe UHT (0.53), mleko niskotłuszczowe (0.77).

Klasyfikujemy w oparciu o jedynie **etykiety produktów** (czyli na podstawie opisu nadanego przez sieć). Klasyfikacja odbywa się po wyselekcjonowaniu fraz kluczowych lub słów kluczowych charakteryzujących daną podgrupę.

Realizacja: środowisko R i autorska funkcja **select_labels**:

```
library(stringr)
```

```
select_labels<-function(data,include=NULL,must=NULL,exclude=NULL)  
{....}
```

Etap1) **Klasyfikujemy produkty do kategorii MLEKO:**

```
mleko<-select_labels(data=MyData, include=c("mleko","Mleko", "Kozie", "kozie", "UHT", "uht", "wydojone"),  
exclude=c("zsiad","Zsiad","czekoladowe", "Czekoladowe", "CZEKOLADOWE","Kefir", "lanka", "trusk", "czek",  
"Jogurt", "jogurt", "mlecz", "jog", "Jogurt", "Jog", "banan", "mietanka", "mietana", "smakowe",  
"MIETANKA","MIETANA","BEBIKO", "NAN", "SOJOWY", "KOKOSOWY", "serwat",  
"kakao","Serek","serek","Czeko","18%","12%","16%","mikrofiltrow","00403995G", "Kanapka", "Ciasto","kulki",  
"mietana", "sos", "Sos"))
```

Etap 2) **Utworzenie zbiorów danych dla grup elementarnych**

```
mleko_kozie<-select_labels(d, include=c("kozie", "Kozie"),exclude=c("proszku"))
```

```
mleko_proszek<-select_labels(d, include=c("proszku", "PROSZKU", "SZCZONE", "szczone", "szcz", "zag",  
"400g", "granul"))
```

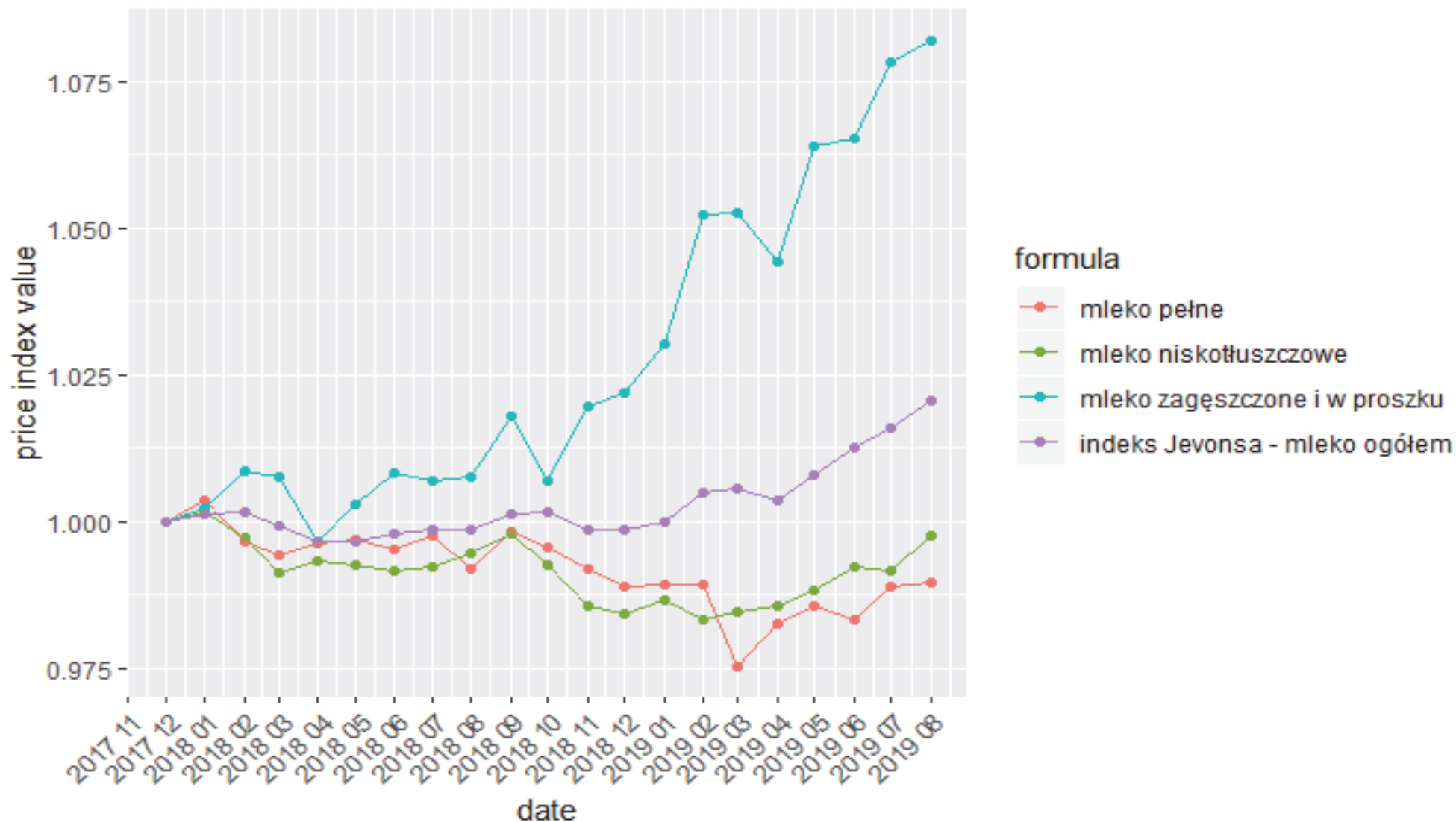
```
mleko_pelne_UHT<-select_labels(d, include=c("3,2%", "3,5%", "3,8%"), must=c("UHT", "uht"),  
exclude=c("0,5%","1,5%","kozie","Kozie", "proszku"))
```

```
mleko_pelne_pasteryzowane<-select_labels(d, include=c("3,2%", "3,5%",  
"3,8%","pasteryz","past","paste"),exclude=c("0,5%","1,5%","kozie","Kozie", "proszku","UHT","uht"))
```

```
mleko_nisko_UHT<-select_labels(d, include=c("0%", "0,5%", "1,5%", "2%", "2,5%"), must=c("UHT", "uht"),  
exclude=c("3,2%","3,5%","kozie","Kozie", "proszku"))
```

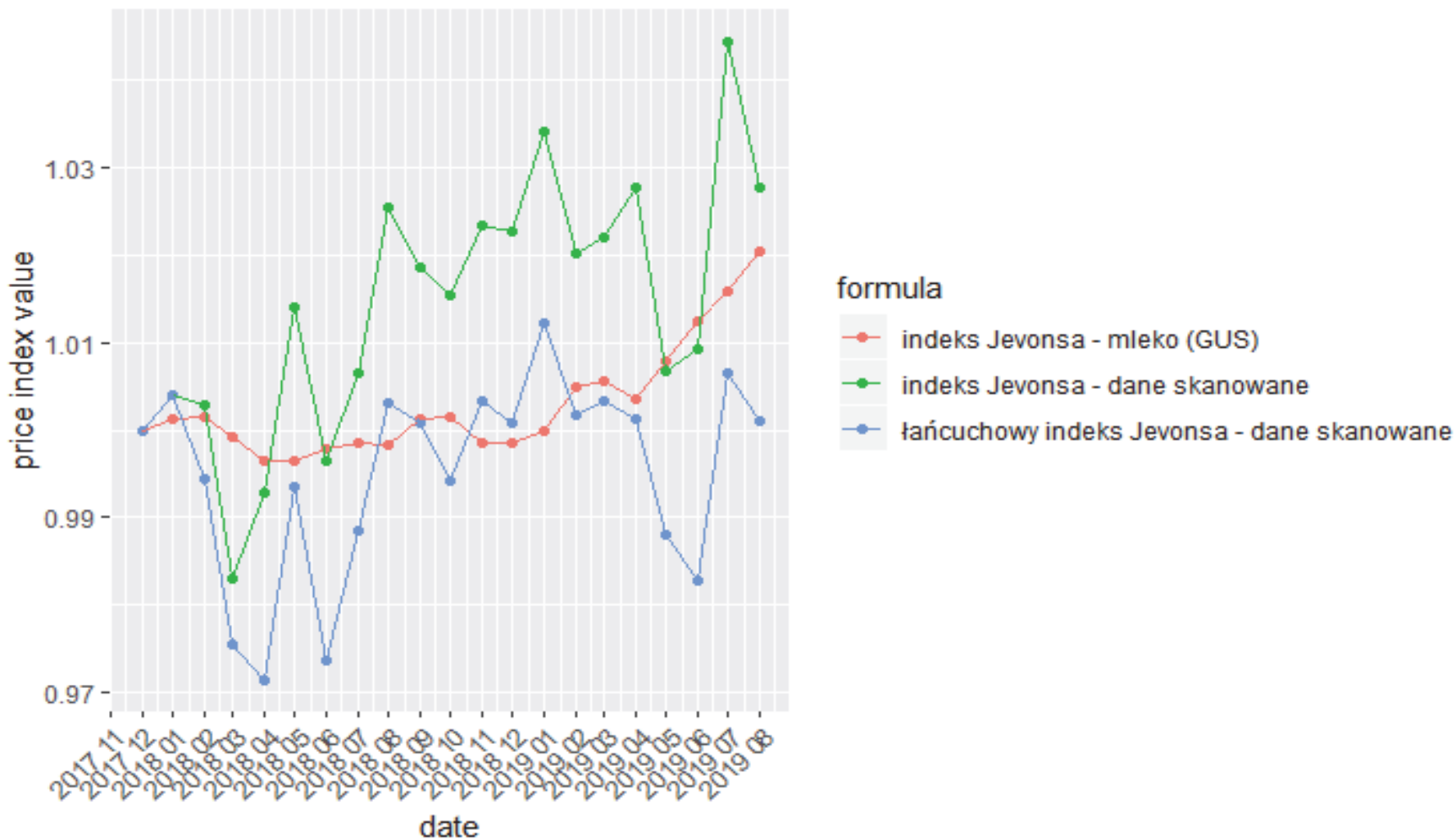
```
mleko_nisko_pasteryzowane<-select_labels(d, include=c("0%", "0,5%", "1,5%", "2%", "2,5%"),  
exclude=c("3,2%","3,5%","kozie","Kozie", "proszku", "UHT", "uht"))
```

Rys. 1. Dynamika cen mleka przy tradycyjnej kolekcji danych (okres referencyjny: grudzień 2017)

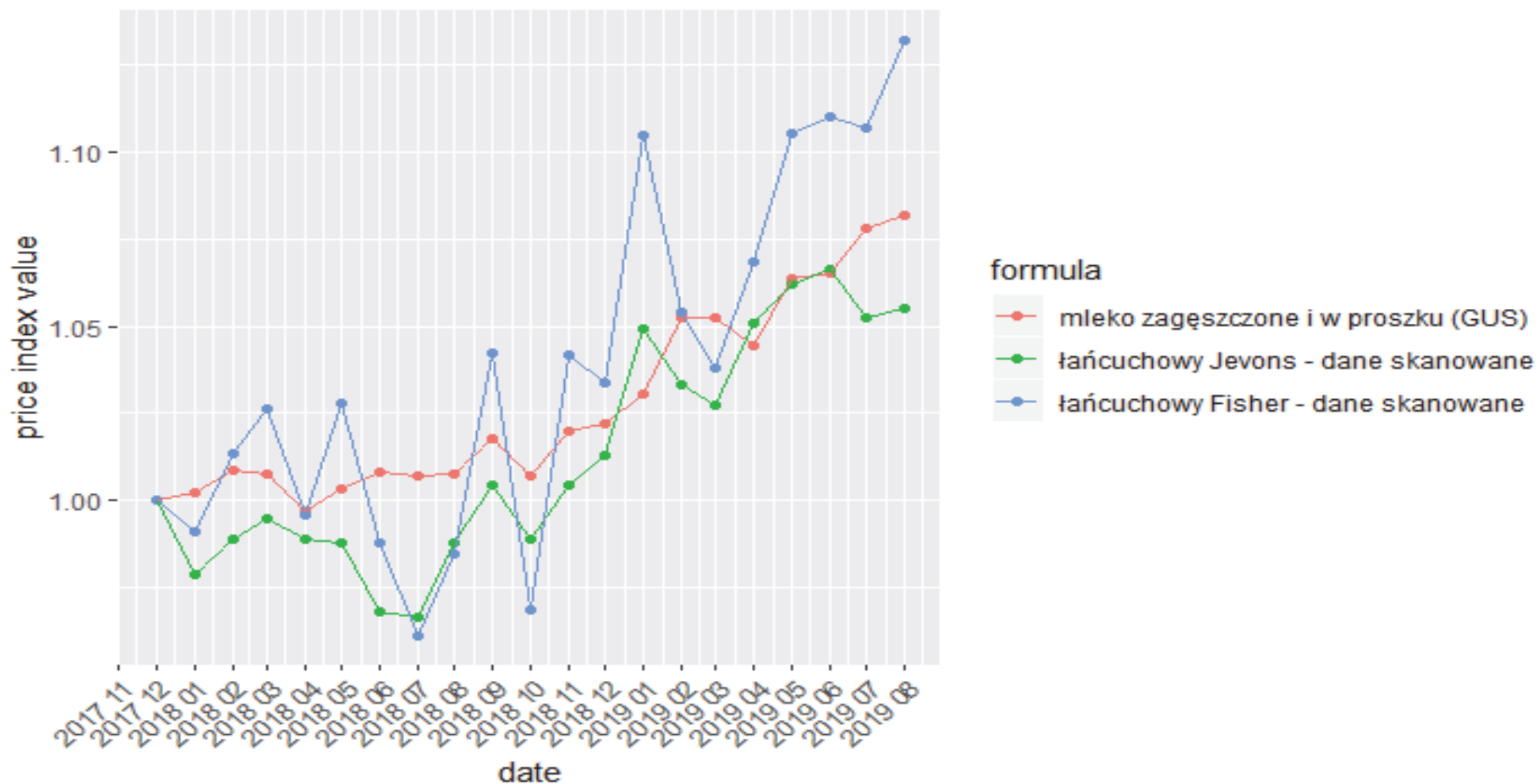


Rys. 2A. Dynamika cen mleka ogółem

(tradycyjna kolekcja danych vs dane skanowane, okres referencyjny: grudzień 2017)

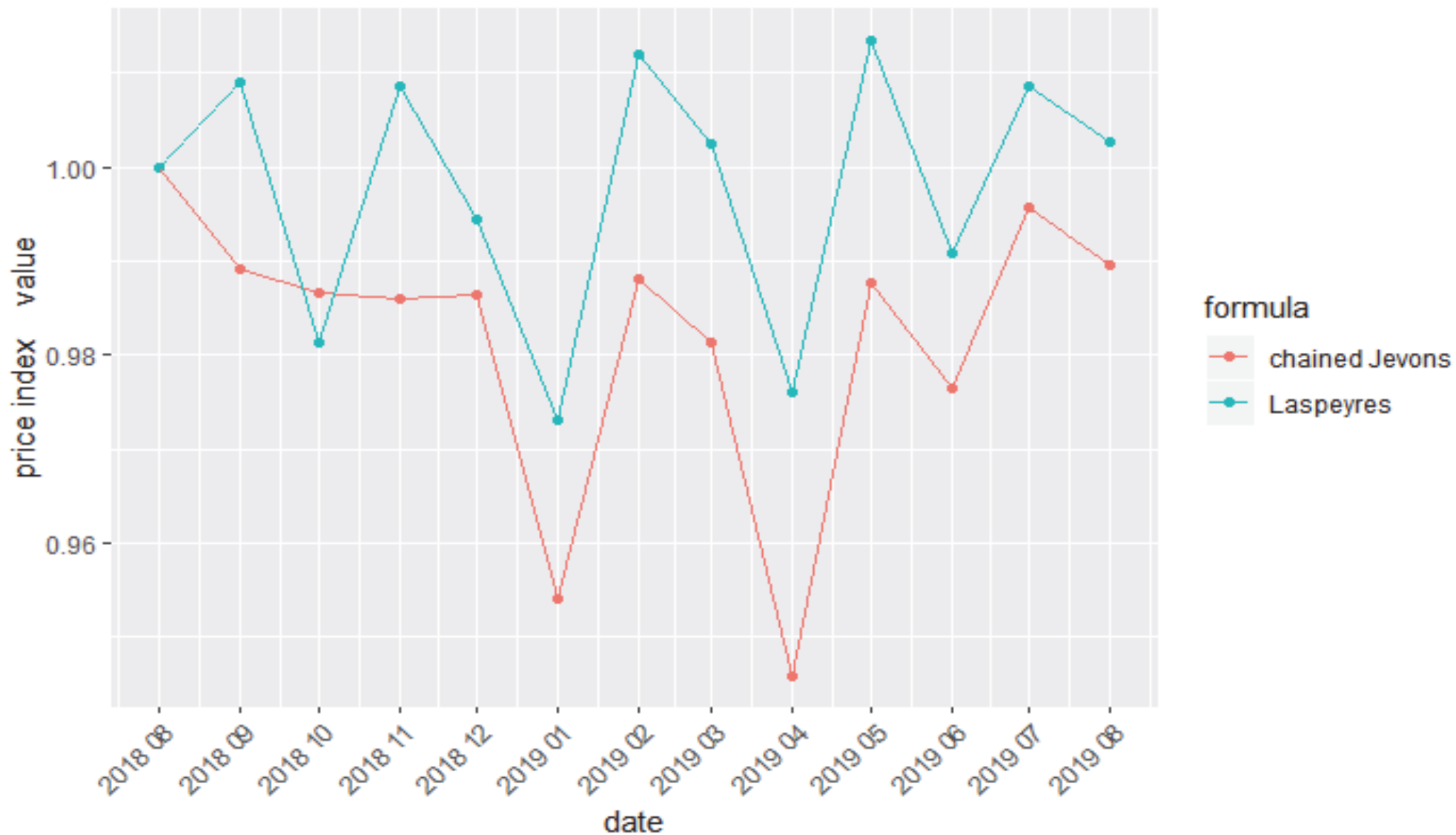


Rys. 2B. Dynamika cen mleka zagęszczonego i w proszku
(tradycyjna kolekcja danych vs dane skanowane, okres referencyjny: grudzień 2017)

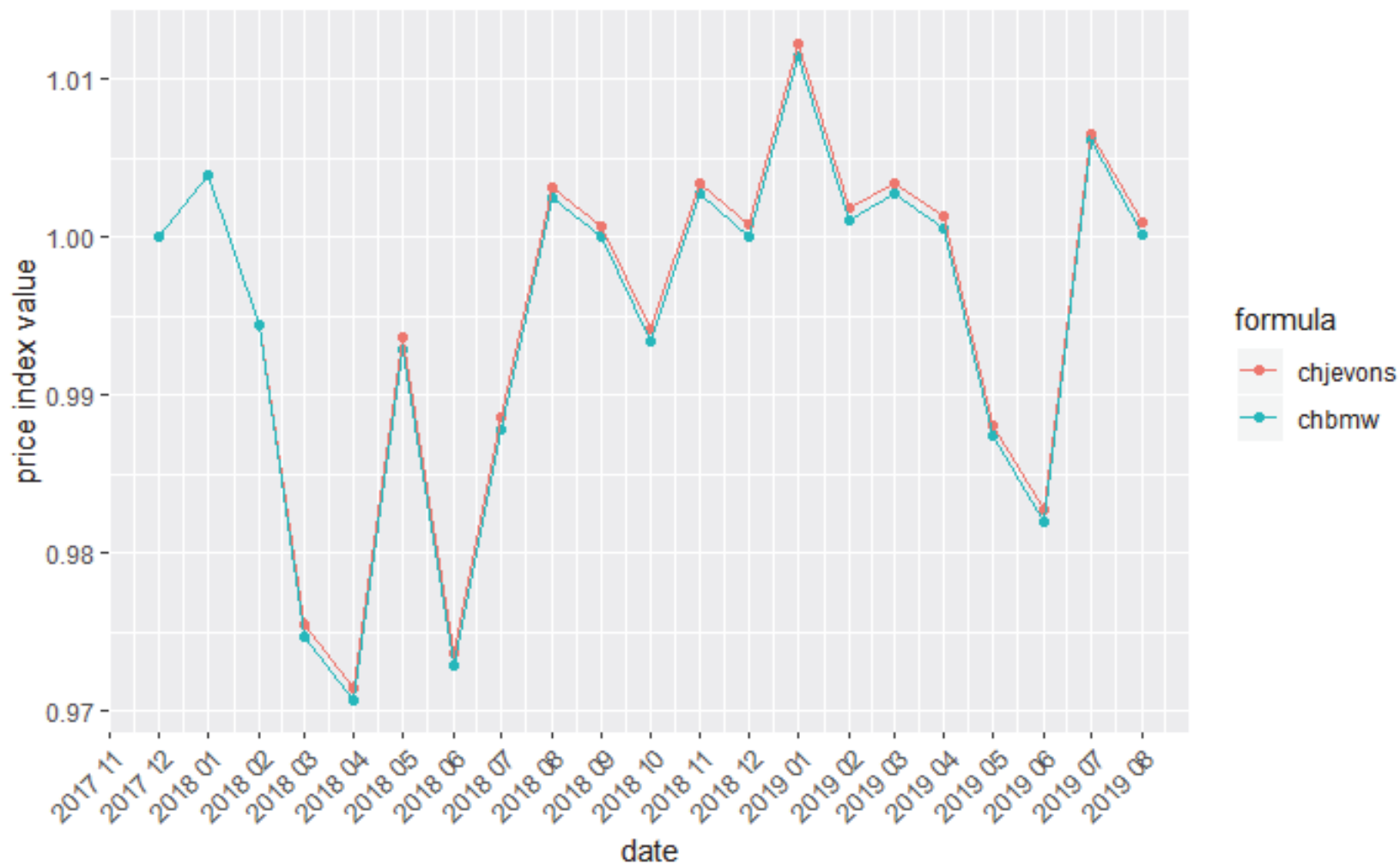


Rys. 3A. Dynamika cen mleka ogółem

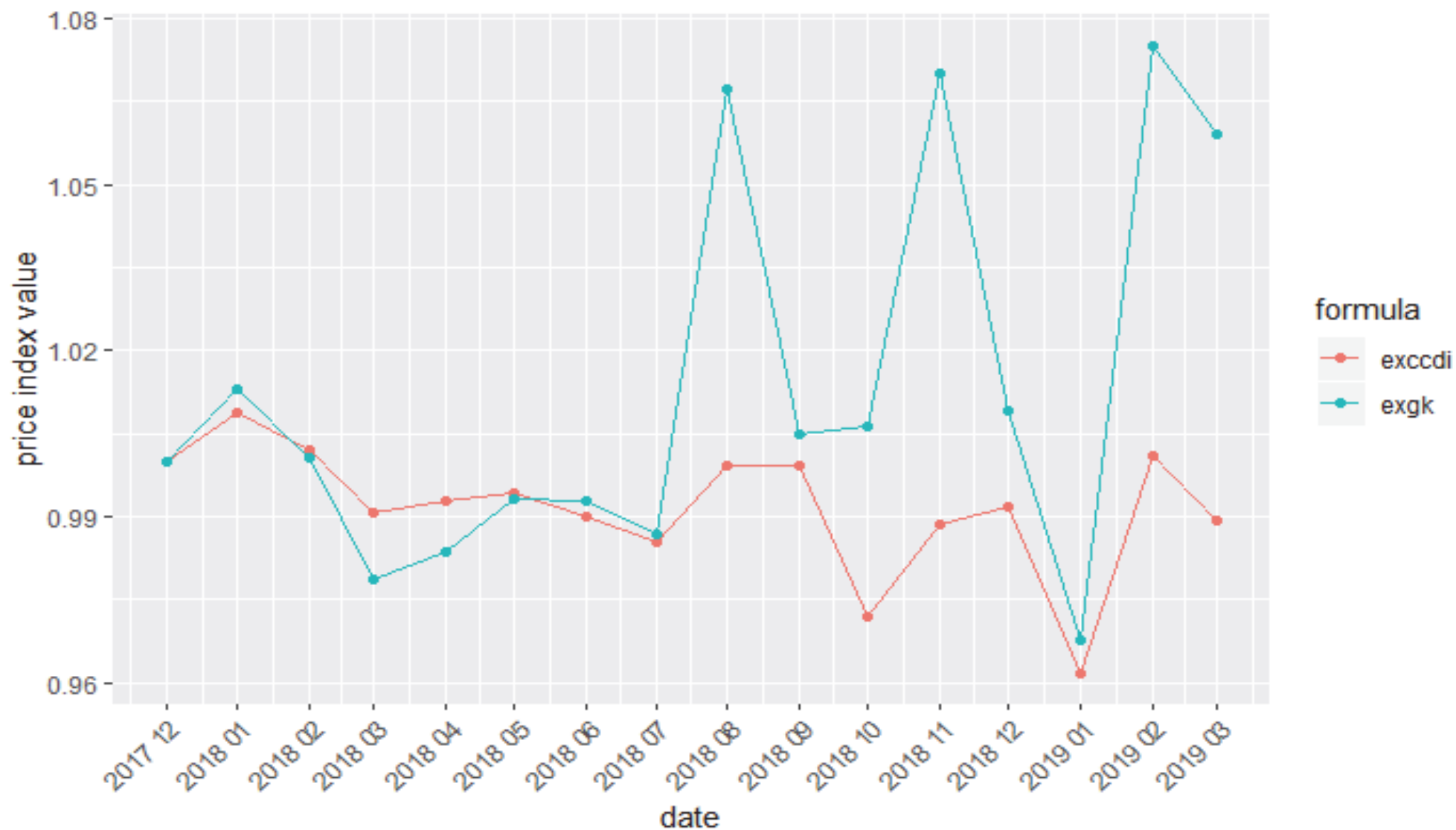
(dane skanowane, okres referencyjny: grudzień 2018)



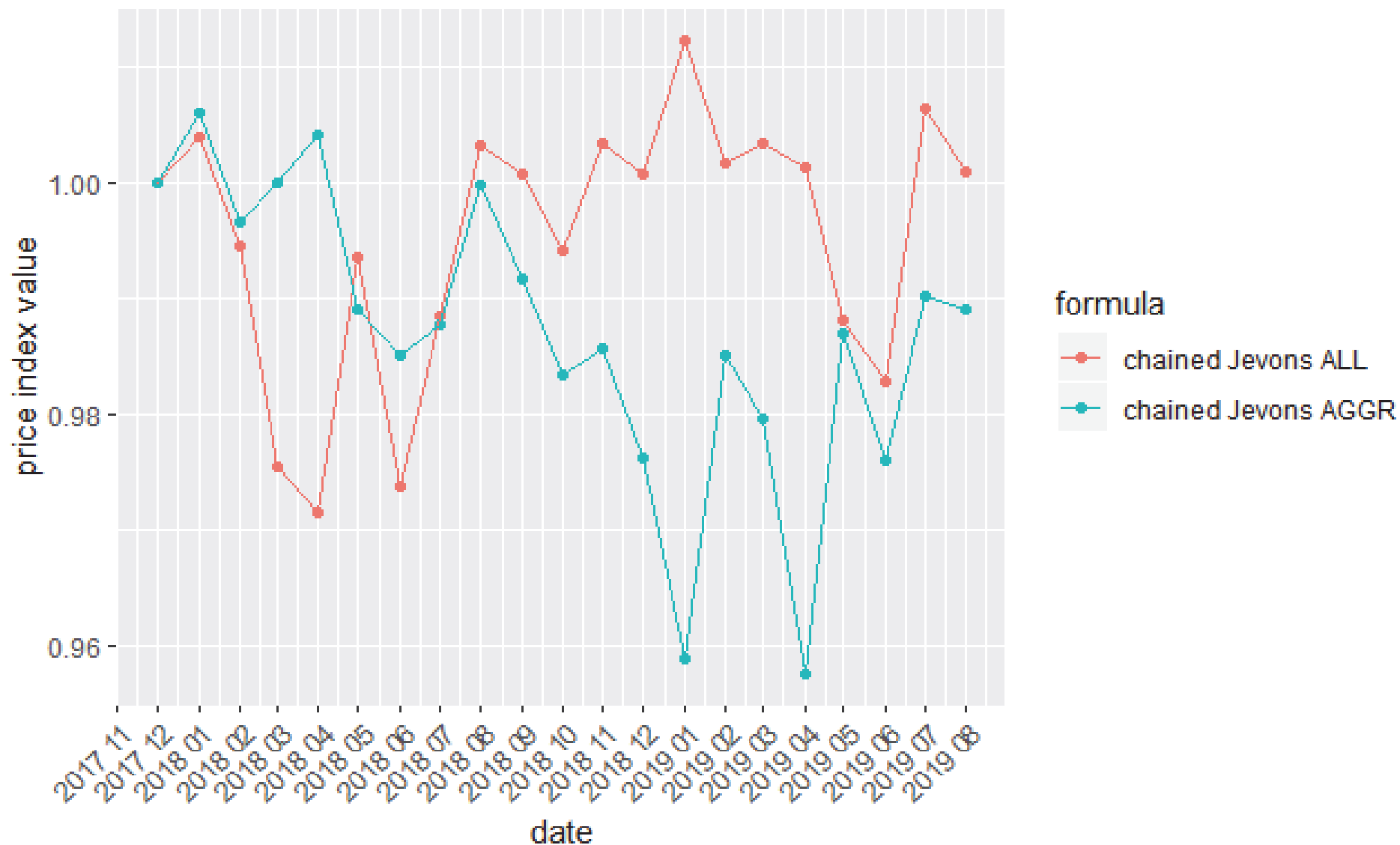
**Rys. 3B. Dynamika cen mleka ogółem –indeksy elementarne
(dane skanowane, okres referencyjny: grudzień 2017)**



Rys. 3C. Dynamika cen mleka ogółem – indeksy multilateralne
(dane skanowane, okres referencyjny: grudzień 2017)

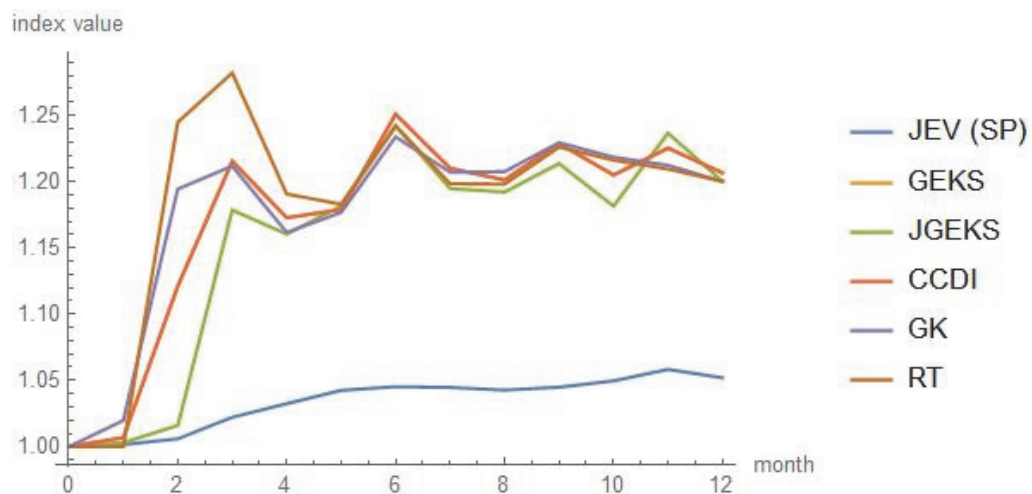


Rys. 4. Dynamika cen mleka (podwójna agregacja vs brak agregacji)
(dane skanowane, okres referencyjny: grudzień 2017)

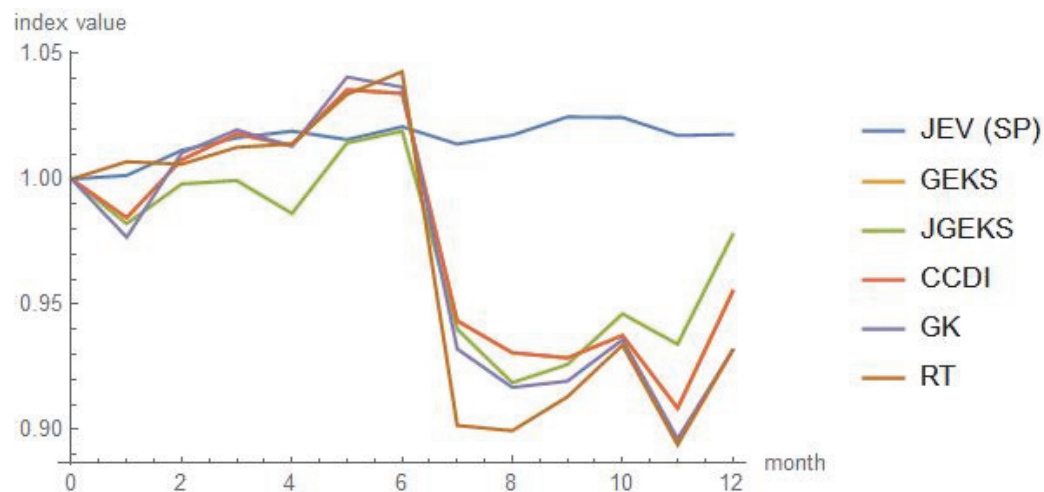


Rys. 5. Dane z **allegro** vs tradycyjna kolekcja
(grupa: rowery, rok analizy: 2018, dane miesięczne)

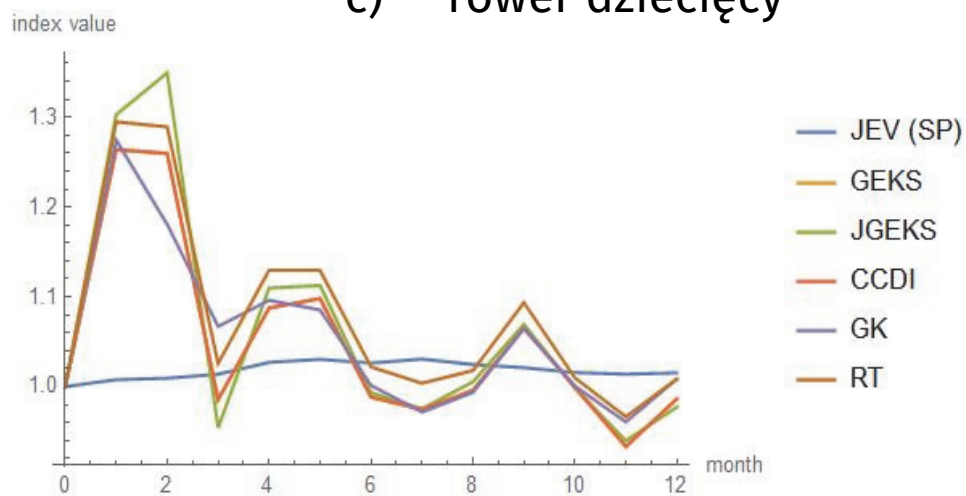
a) rower górski



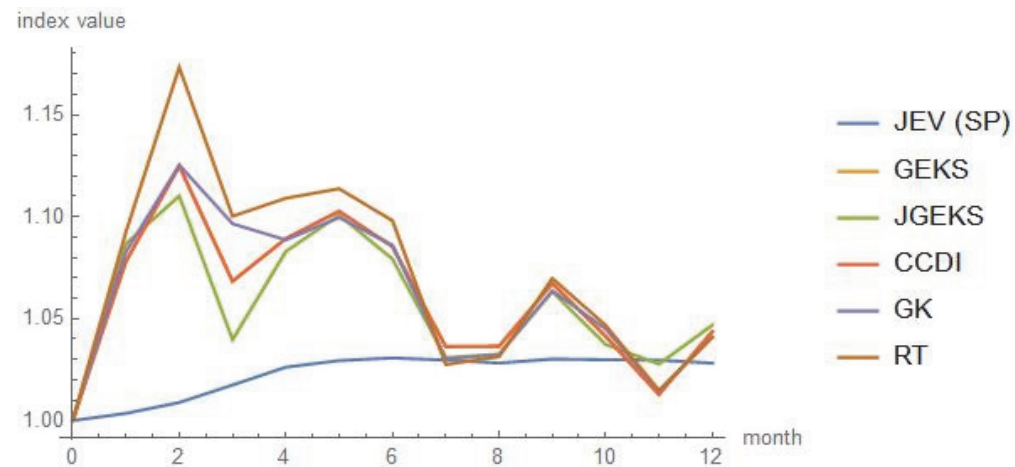
b) rower trekkingowy



c) rower dziecięcy



d) rowery ogółem



Dziękuję za uwagę!

BIBLIOGRAFIA

- Białek, J., Bobel, A. (2019). Comparison of Price Index Methods for CPI Measurement using Scanner Data, *Paper presented at the 16th Meeting of the Ottawa Group on Price Indices*. Rio de Janeiro, Brazil.
- Białek J., Roszko-Wójtowicz (2019), The Impact of the Price Index Formula on the Consumer Price Index Measurement, *Statistika – Statistics and Economy Journal*, Vol. 99 (3), 246-258, Czech Statistical Office, Praga.
- Caves D.W., Christensen, L.R., Diewert, W.E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal*, 92, 73-86.
- Chessa, A. G. (2015). Towards a generic price index method for scanner data in the Dutch CPI, *Room document for Ottawa Group Meeting*. Urayasu City, Japan.
- Chessa, A.G. (2016). A New Methodology for Processing Scanner Data in the Dutch CPI. *Eurona*, 2016(1), 49-69.
- Chessa, A.G. (2017). Comparisons of QU-GK Indices for Different Lengths of the Time Window and Updating Methods, *Paper prepared for the second meeting on multilateral methods organised by Eurostat*. Luxembourg, Statistics Netherlands.
- Chessa, A.G., Verburg, J., Willenborg, L. (2017). A comparison of price index methods for scanner data, *Paper presented at the 15th Meeting of the Ottawa Group on Price Indices*. Eltville am Rhein, Germany.
- Chessa, A. G. (2018). Product definition and index calculation with MARS-QU: Applications to consumer electronics. *Report Statistics Netherlands*
- Consumer Price Index Manual. Theory and practice. (2004). International Labour Office (ILO), Geneva.
- Dalen, J. (1997). Experiments with Swedish Scanner Data. *Proceedings of the Third Meeting of the International Working Group on Price Indexes*, Balk (ed.), Research Paper no. 9806, Statistics Netherlands, Division Research and Development, Department of Statistical Methods.
- Dalen, J. (2017). Unit values in scanner data and some operational issues, *Paper presented at the fifteenth Ottawa Group Meeting*. Eltville am Rhein, Germany.
- Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4, 114-145.
- Diewert, W.E., Fox, K.J. (2017). Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data. *Discussion paper*, 17(2), Vancouver School of Economics, The University of British Columbia, Vancouver, Canada.
- Eltető, Ö., Köves, P. (1964). On a Problem of Index Number Computation Relating to International Comparisons/ (in Hungarian). *Statisztikai Szemle*, 42, 507-518.
- Fisher, I. (1922). *The Making of Index Numbers*. Boston: Houghton Mifflin.
- Geary, R.G. (1958). A Note on Comparisons of Exchange Rates and Purchasing Power between Countries. *Journal of the Royal Statistical Society Series A*, 121, 97-99.
- Guerreiro, V., Walzer, M., Lamboray, C. (2018), The use of Supermarket Scanner data in the Luxemburg Consumer Price Index. *Working papers du STATEC, Economie et Statistiques* 97,1-18.
- Gini, C. (1931). On the Circular Test of Index Numbers. *Metron*, 9:9, 3-24.
- Inklaar, R., Diewert, W. E. (2016). Measuring Industry Productivity and Cross-Country Convergence. *Journal of Econometrics*, 191, 426-433.
- Jevons, W.S. (1865). The variation of prices and the value of the currency since 1782. *J. Statist. Soc. Lond.*, 28, 294-320.
- Khamis, S.H. (1972). A New System of Index Numbers for National and International Purposes. *Journal of the Royal Statistical Society Series A*, 135, 96-121.
- Krsinich, F. (2014). The FEWS Index: Fixed Effects with a Window Splice – Non-Revisable Quality-Adjusted Price Indices with No Characteristic Information, *Paper presented at the meeting of the group of experts on consumer price indices* (s. 26-28). Geneva, Switzerland.
- Laspeyres, E. (1871). Die Berechnung einer mittleren Waarenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik*, 16, 296–314.
- Leonard, I., Sillard, P., Varlet, G. Zoyem, J. P. (2017). Scanner data and quality adjustment. *Serie des Documents de Travail*, Working Paper No. F1704, INSEE, 1-31.
- Loon, K. V., Roels, D. (2018). Integrating big data in the Belgian CPI, *Paper presented at the meeting of the group of experts on consumer price indices* (s. 8-9). Geneva, Switzerland.
- Maddison, A., Rao, D.S.P. (1996). A Generalized Approach to International Comparison of Agricultural Output and Productivity, Research memorandum GD-27. *Groningen Growth and Development Centre*, Groningen, The Netherlands.
- Paasche, H. (1874). Über die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen. *Jahrbücher für Nationalökonomie und Statistik*, 12, 168–178.
- Saraiva dos Santos, P., Lidonio, F., Cardoso, C. (2012). Scanner Data Project: the experience of Statistics Portugal, *Paper presented at the Workshop on Scanner Data* (s. 1-13), Stockholm.
- Szulc, B. (1964). Indices for Multiregional Comparisons. (in Polish). *Przegląd Statystyczny*, 3, 239-254.

SEMINARIUM

Pomiar inflacji

Stan obecny i wyzwania

