Urząd Statystyczny
w Poznaniu

Imputacja danych dotyczących wynagrodzenia w Badaniu Aktywności Ekonomicznej Ludności

Kamil Wilak[1,2], Andrzej Młodak[1,3], Tomasz Klimanek[1,2], Tomasz Józefowski[1,2]

[1]Urząd Statystyczny w Poznaniu
[2]Uniwersytet Ekonomiczny w Poznaniu
[3]Akademia Kaliska im. Prezydenta Stanisława Wojciechowskiego

MET 2023, 3-5 lipca 2023

Plan prezentacji

1. Wprowadzenie
2. Problem braków odpowiedzi w wynagrodzeniach w BAEL
3. Analiza wynagrodzeń w BAEL
4. Propozycja procedury imputacji wynagrodzenia w BAEL
5. Dalsze kierunki badań

Wprowadzenie

Projekt

Obszar statystyki pracy - Infrastruktura statystyczna LFS w ramach zintegrowanej europejskiej statystyki społecznej (IESS) – Moduł 2023; Pakiet roboczy 2. Imputacja zmiennej „miesięczne wynagrodzenie z pracy głównej”

LABOUR DOMAIN - LFS STATISTICAL INFRASTRUCTURE UNDER IESS - LFS 2023 MODUL; WORK PACKAGE 2. IMPUTATION OF DATA FOR INFORMATION ON INCOME FROM WORK

101101153 — 2022-PL-LFS-LMB

Kierownik Projektu

Hanna Strzelecka

Jednostki realizujące

- Główny Urząd Statystyczny
- US Poznań
- US Łódź

Termin realizacji projektu

2022.09.01–2024.05.31

Wprowadzenie

Cel badania

Wypracowanie metody imputacji zmiennych dotyczących wynagrodzenia netto w Badaniu Aktywności Ekonomicznej Ludności (BAEL), pozwalającej na uzyskanie danych o akceptowalnej jakości.

Dane

Dane jednostkowe BAEL z kwartałów: 2021 I – 2022 II

Problem braków odpowiedzi w wynagrodzeniach w BAEL

Odsetek odpowiedzi na pytanie o wynagrodzenie netto w kwartalnych próbach, 2021 Q1 – 2022 Q2

	Kwartał					
	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2
ogółem						
	26.1	24.8	21.7	19.3	18.5	18.0
województwo						
dolnośląskie	29.6	31.8	27.6	23.8	25.2	26.5
kujawsko-pomorskie	31.6	30.2	26.6	24.1	22.6	22.8
lubelskie	33.4	32.7	32.9	31.5	28.3	23.7
lubuskie	25.5	21.9	20.3	17.1	17.8	19.0
łódzkie	12.8	12.3	13.2	9.4	7.8	6.9
małopolskie	19.6	18.3	16.5	14.1	14.2	16.0
mazowieckie	37.2	34.0	26.6	22.5	23.0	23.0
opolskie	13.5	10.9	5.9	8.2	12.8	10.7
podkarpackie	10.4	9.8	9.3	8.7	5.8	4.9
podlaskie	33.3	34.8	31.6	28.2	25.3	25.1
pomorskie	27.6	26.0	24.3	21.5	18.8	19.1
śląskie	26.9	25.7	23.0	21.5	20.0	19.4
świętokrzyskie	26.2	27.7	22.6	17.9	19.6	22.9
warmińsko-mazurskie	32.5	24.8	23.6	22.9	19.2	14.5
wielkopolskie	23.8	21.0	16.2	13.9	13.2	12.0
zachodniopomorskie	17.8	18.4	17.5	15.4	15.0	12.8

Źródło: Opracowanie własne na podstawie BAEL

Problem braków odpowiedzi w wynagrodzeniach w BAEL

Odsetek odpowiedzi na pytanie o przedział wynagrodzenia netto w kwartalnych próbach, 2021 Q1 – 2022 Q2

	Kwartał					
	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2
ogółem						
	16.5	14.5	14.5	13.8	16.8	18.2
województwo						
dolnośląskie	20.0	16.3	15.0	17.4	18.2	18.2
kujawsko-pomorskie	20.5	16.8	14.5	11.8	13.4	13.0
lubelskie	15.2	14.2	16.1	15.2	16.9	19.6
lubuskie	15.9	9.9	9.7	7.5	10.6	13.3
łódzkie	17.6	12.5	14.3	9.3	8.1	8.4
małopolskie	20.3	17.9	18.9	15.8	24.5	20.3
mazowieckie	14.3	13.4	13.2	14.6	20.8	21.3
opolskie	9.9	5.2	4.3	7.4	11.7	12.0
podkarpackie	11.1	11.6	8.7	10.8	13.1	18.3
podlaskie	24.5	20.9	23.3	18.1	19.7	20.8
pomorskie	16.5	17.0	14.9	13.1	15.2	16.2
śląskie	14.7	13.8	12.8	12.8	18.2	24.5
świętokrzyskie	18.1	16.1	15.7	17.5	20.6	20.3
warmińsko-mazurskie	18.8	22.4	27.1	26.0	23.5	24.4
wielkopolskie	14.6	11.9	13.8	11.3	12.5	15.3
zachodniopomorskie	16.8	14.0	11.6	12.9	15.5	14.3

Źródło: Opracowanie własne na podstawie BAEL

Problem braków odpowiedzi w wynagrodzeniach w BAEL

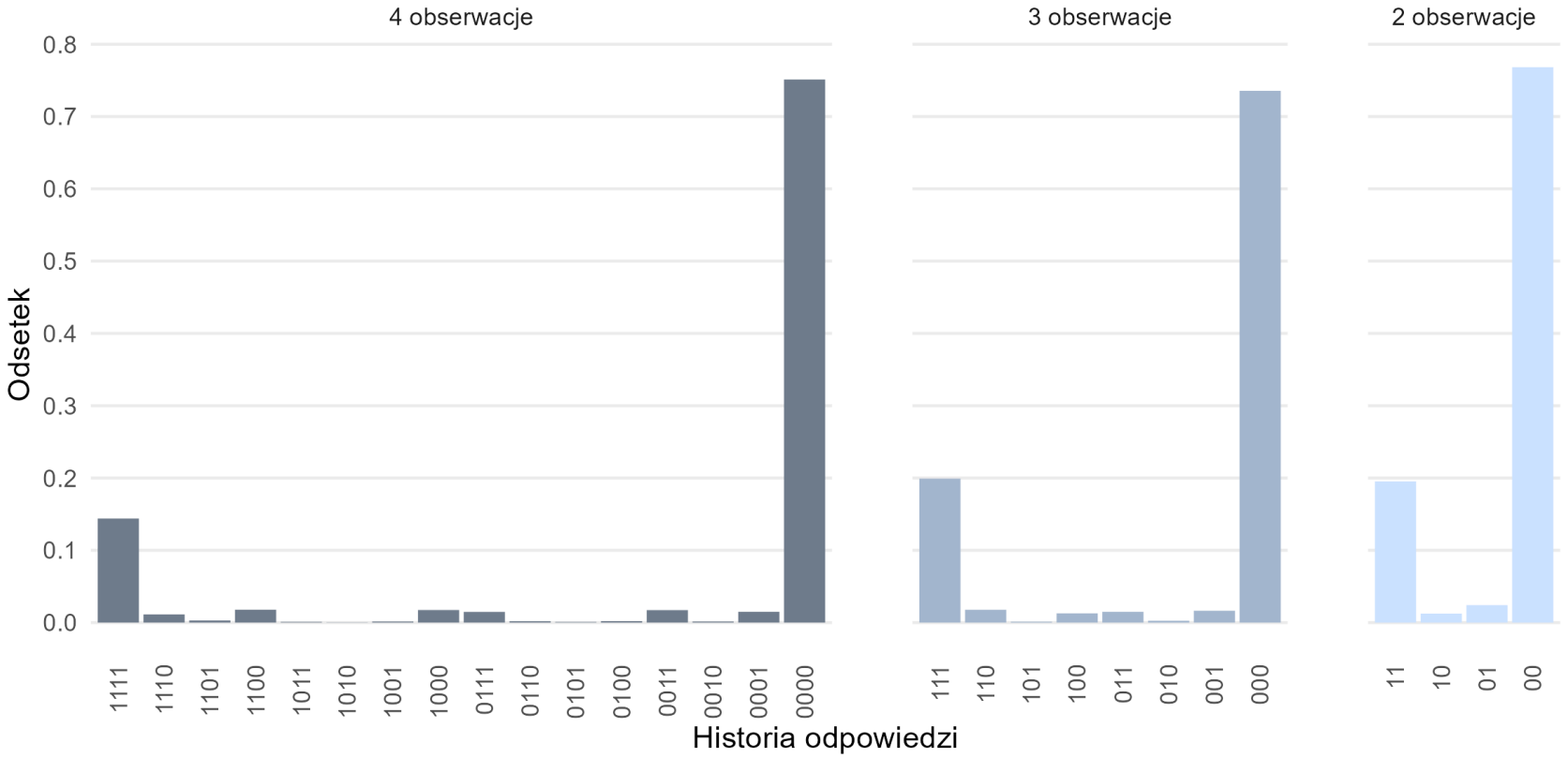
Odsetek odpowiedzi na pytanie o wynagrodzenie netto lub przedział wynagrodzenia netto w kwartalnych próbach, 2021 Q1 – 2022 Q2

	Kwartał					
	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2
ogółem						
	42.6	39.2	36.2	33.1	35.2	36.3
województwo						
dolnośląskie	49.5	48.1	42.6	41.2	43.4	44.8
kujawsko-pomorskie	52.0	46.9	41.1	35.8	36.0	35.9
lubelskie	48.6	47.0	49.0	46.7	45.2	43.2
lubuskie	41.3	31.8	30.0	24.5	28.4	32.3
łódzkie	30.5	24.8	27.5	18.7	15.8	15.3
małopolskie	39.9	36.2	35.4	30.0	38.7	36.4
mazowieckie	51.5	47.4	39.8	37.1	43.8	44.3
opolskie	23.4	16.1	10.3	15.6	24.5	22.7
podkarpackie	21.5	21.4	18.0	19.5	18.9	23.2
podlaskie	57.8	55.7	54.9	46.3	45.0	45.9
pomorskie	44.1	43.0	39.2	34.6	34.0	35.3
śląskie	41.6	39.5	35.8	34.3	38.2	43.9
świętokrzyskie	44.2	43.8	38.3	35.4	40.1	43.1
warmińsko-mazurskie	51.3	47.2	50.7	48.9	42.8	38.9
wielkopolskie	38.4	32.9	29.9	25.2	25.7	27.2
zachodniopomorskie	34.6	32.4	29.1	28.3	30.5	27.0

Źródło: Opracowanie własne na podstawie BAEL

Problem braków odpowiedzi w wynagrodzeniach w BAEL

Udzielanie odpowiedzi na pytanie o wynagrodzenie netto w kolejnych realizacjach badania, 2021 Q1 – 2022 Q2

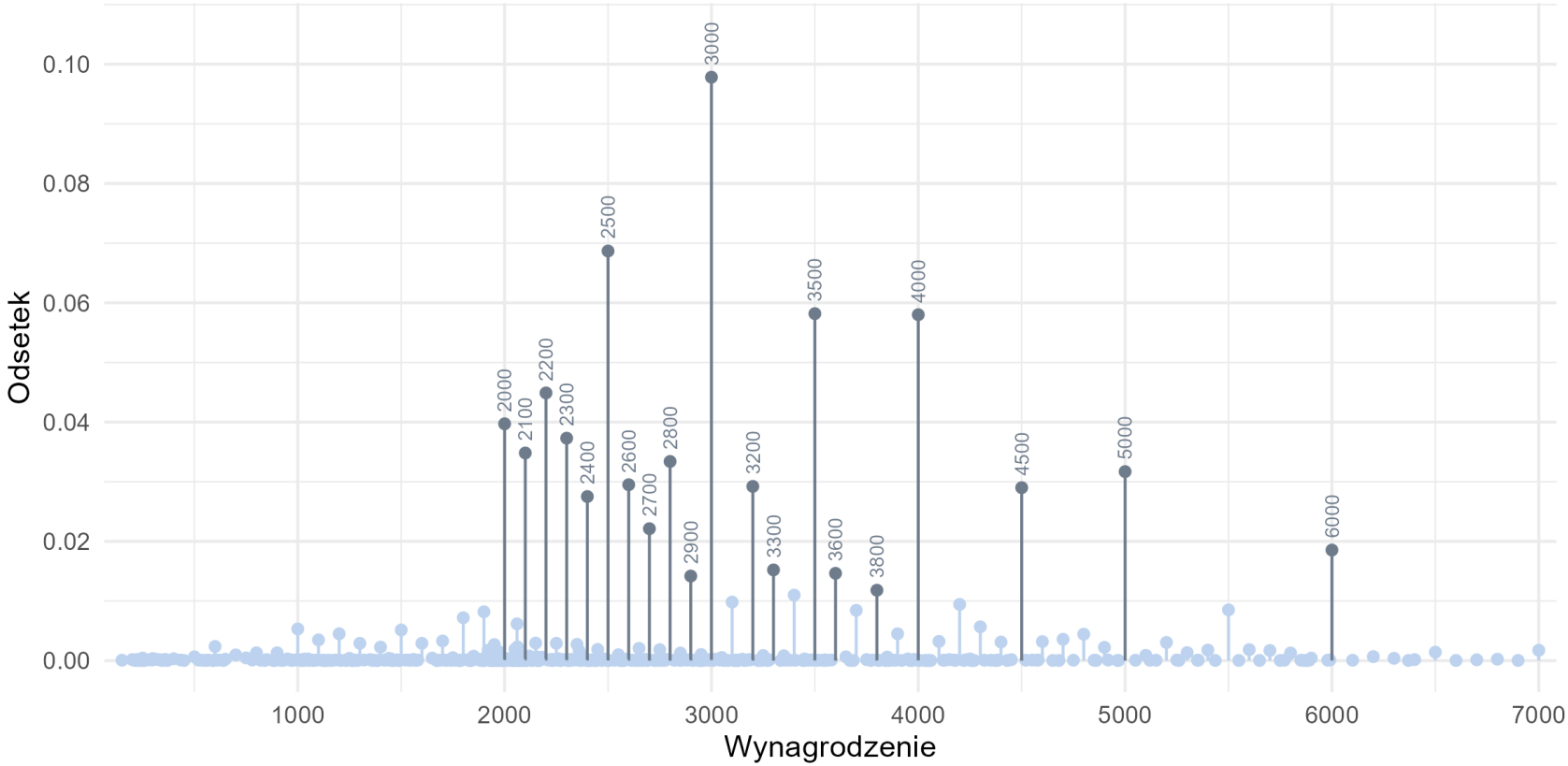


Źródło: Opracowanie własne na podstawie BAEL

Uwaga: Na przykład ciąg 110 dotyczy osób, które trzykrotnie brały udział w badaniu, w przypadku pierwszej i drugiej obserwacji podały wynagrodzenie, zaś w przypadku trzeciej obserwacji – nie

Analiza wynagrodzeń w BAEL

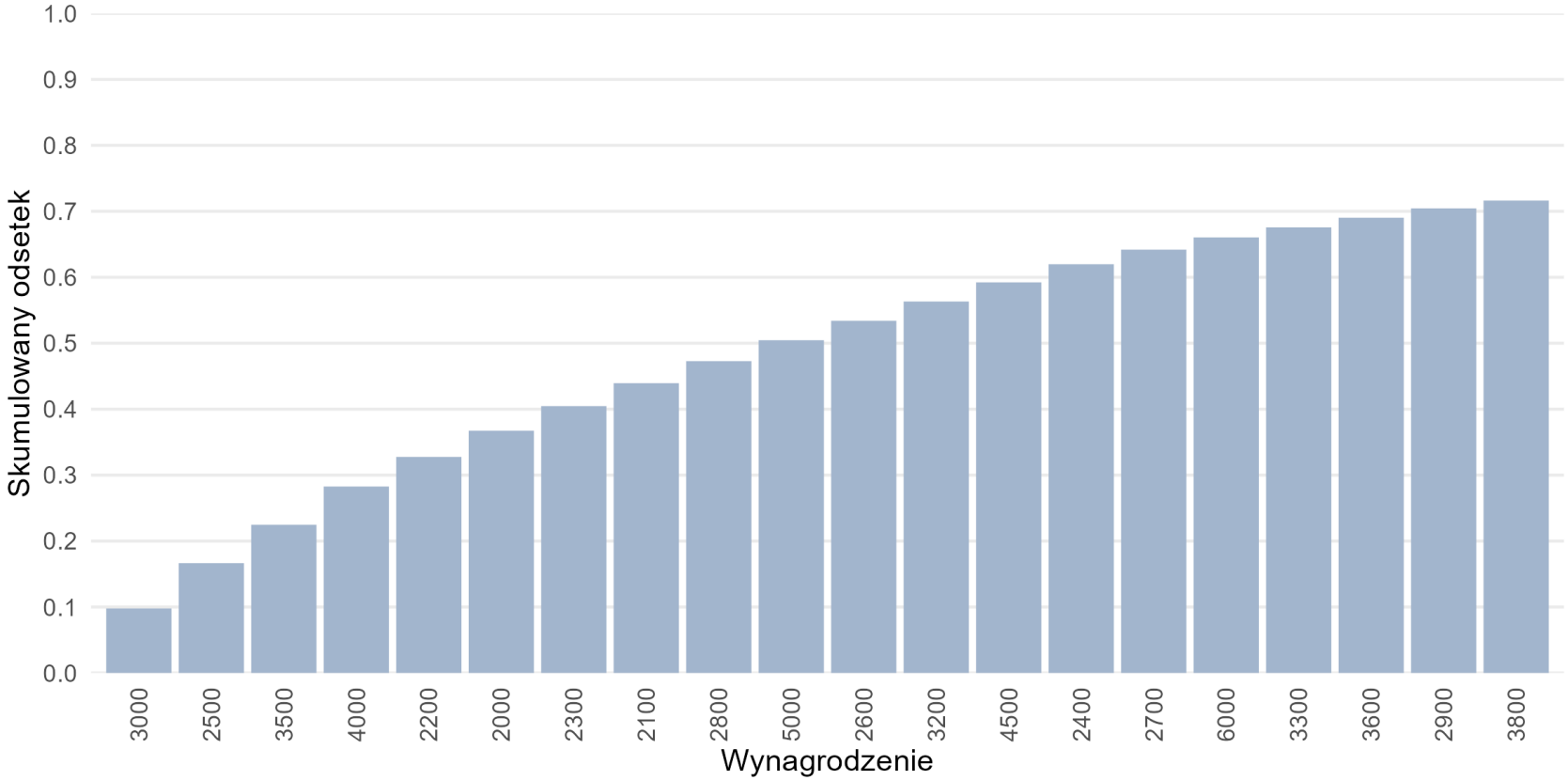
Rozkład wynagrodzeń netto, 2021 Q1 – 2022 Q2



Źródło: Opracowanie własne na podstawie BAEL

Analiza wynagrodzeń w BAEL

Rozkład wynagrodzeń netto, 2021 Q1 – 2022 Q2



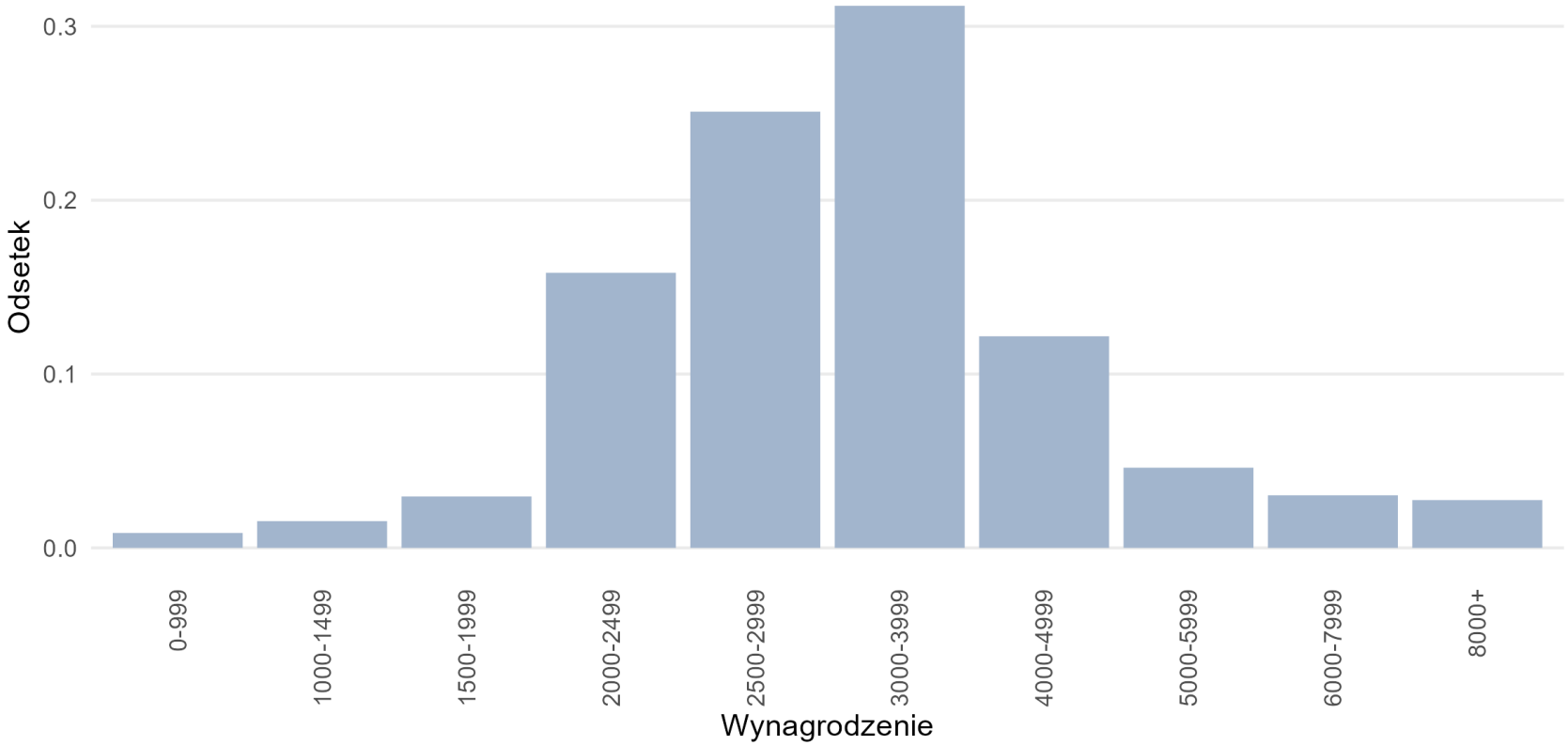
Źródło: Opracowanie własne na podstawie BAEL

Uwaga 1.: Kolejność wynagrodzeń malejąca według liczby wystąpień

Uwaga 2.: Na przykład wysokość słupka dla wynagrodzenia 4000 oznacza, że 27,6% obserwacji wynagrodzenia, stanowiły wartości 3000, 2500, 3500 i 4000

Analiza wynagrodzeń w BAEL

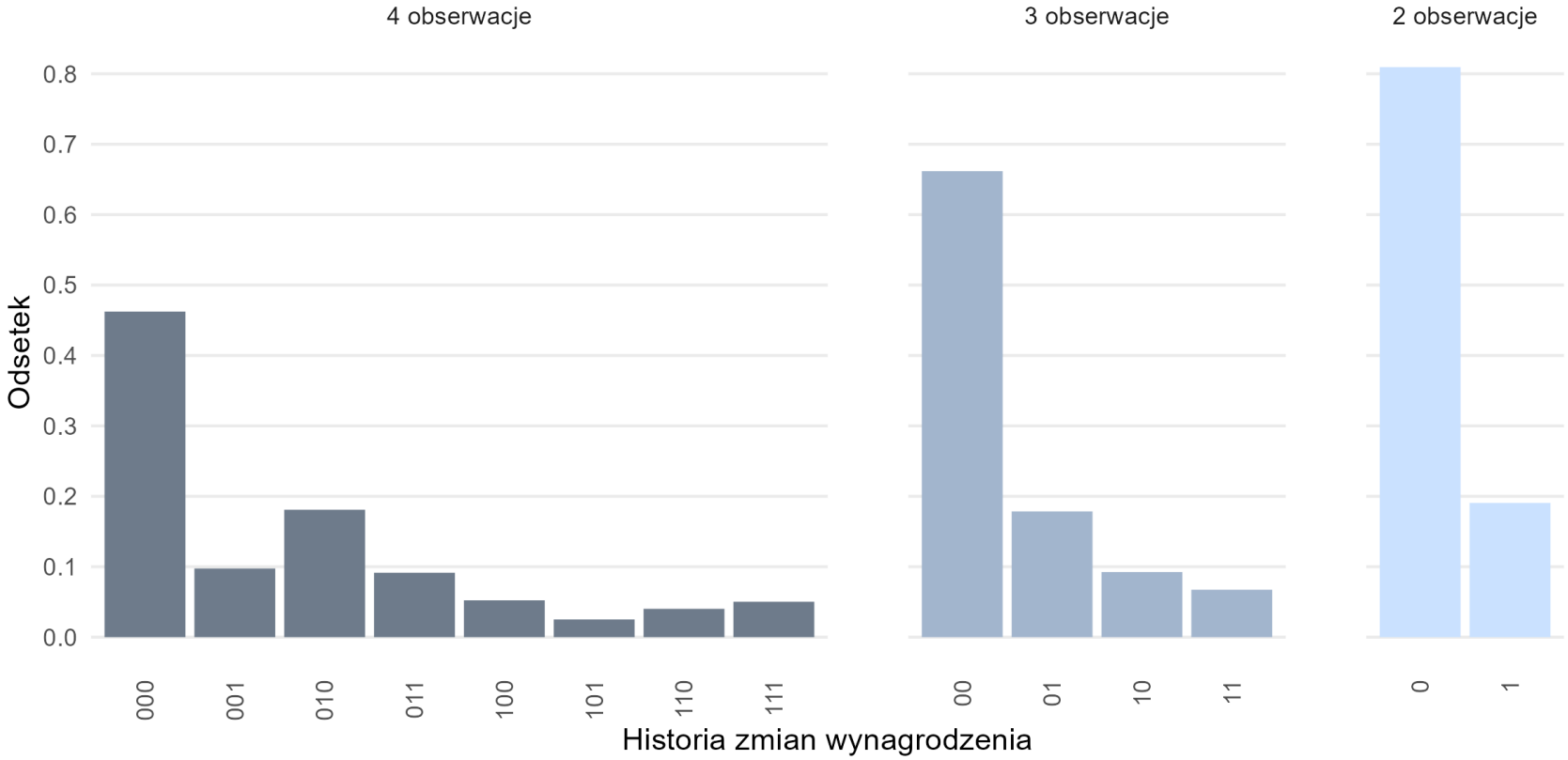
Rozkład odpowiedzi na pytanie o przedział wynagrodzenia netto, 2021 Q1 – 2022 Q2



Źródło: Opracowanie własne na podstawie BAEL

Analiza wynagrodzeń w BAEL

Zmiany wynagrodzeń netto w kolejnych realizacjach badania, 2021 Q1 – 2022 Q2



Źródło: Opracowanie własne na podstawie BAEL

Uwaga: Na przykład ciąg 001 dotyczy osób czterokrotnie pytanych o wynagrodzenie (więc mogły nastąpić potencjalnie trzy zmiany wynagrodzenia), których w drugiej i trzeciej obserwacji wynagrodzenie nie zmieniało się, natomiast w czwartej obserwacji nastąpiła zmiana

Analiza wynagrodzeń w BAEL

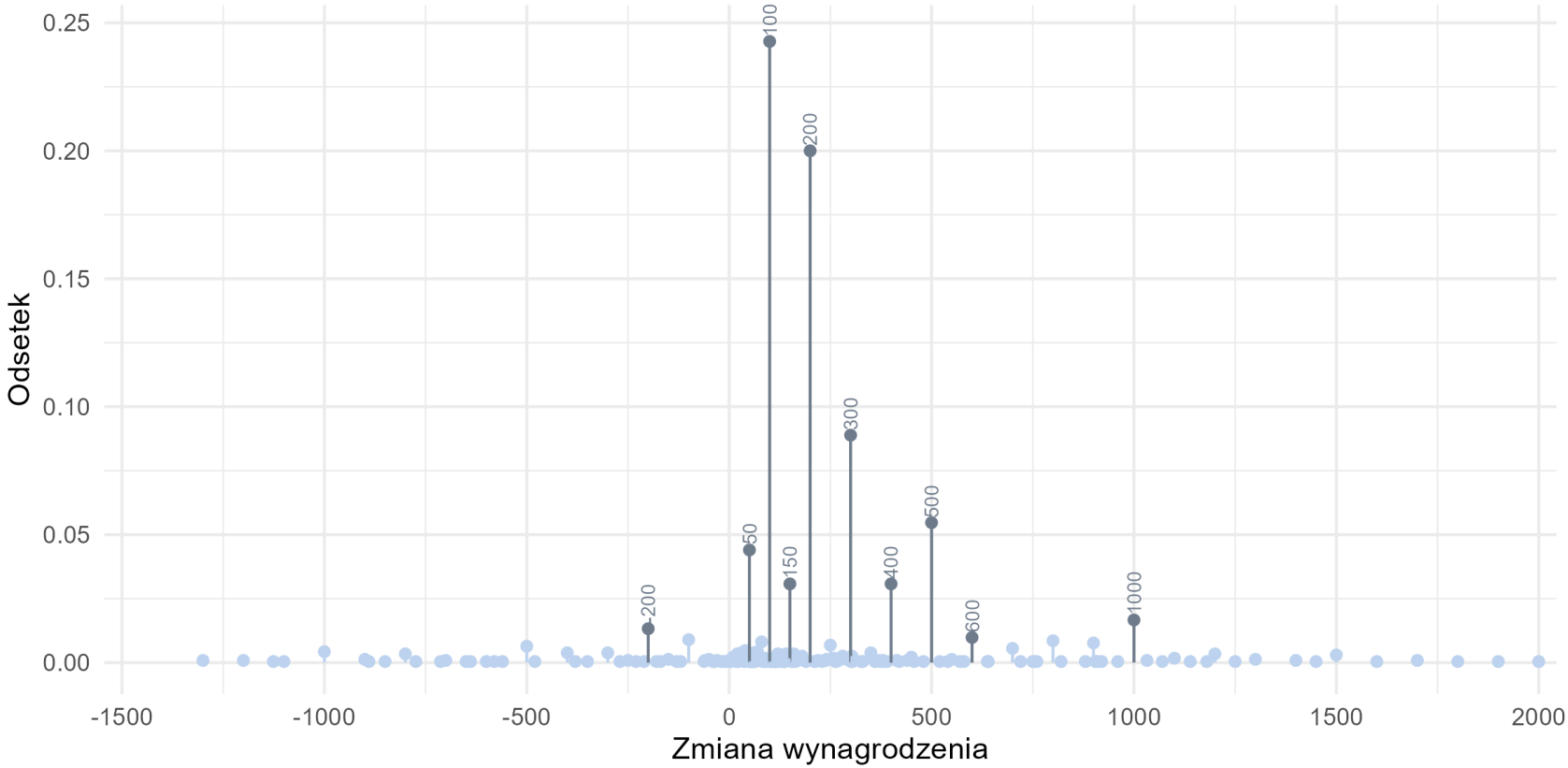
Zmiany wynagrodzeń netto pomiędzy kolejnymi kwartałami, 2021 Q1 – 2022 Q2

	średnia (zł)	minus (%)	zero (%)	plus (%)
2021 Q1 - 2021 Q2	31,19	1,6	83,9	14,5
2021 Q2 - 2021 Q3	22,89	1,5	87,0	11,6
2021 Q3 - 2021 Q4	22,95	1,5	86,8	11,6
2021 Q4 - 2022 Q1	70,42	2,3	68,8	28,9
2022 Q1 - 2022 Q2	42,38	1,7	77,1	21,2

Źródło: Opracowanie własne na podstawie BAEL

Analiza wynagrodzeń w BAEL

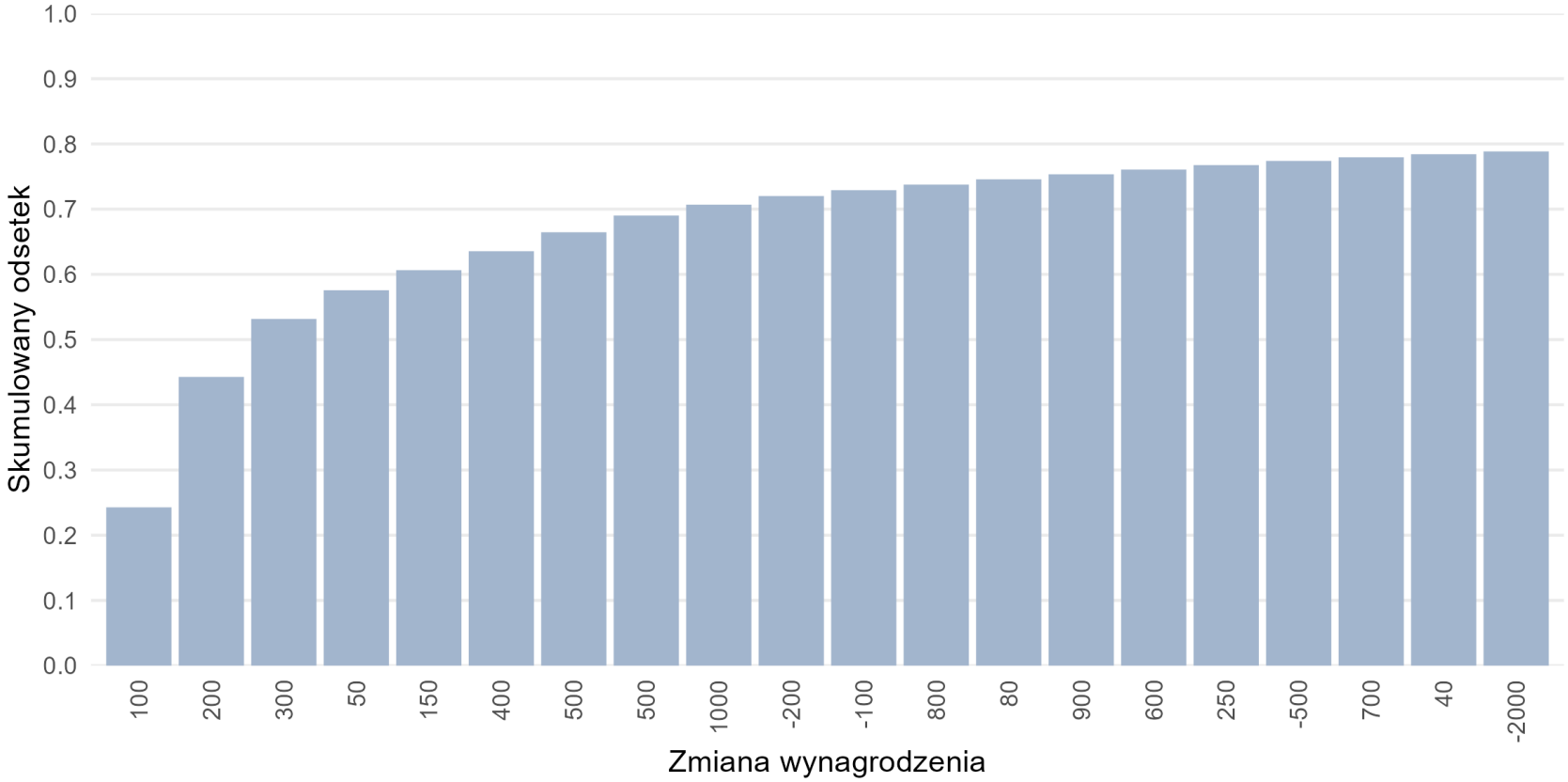
Zmiany wynagrodzeń netto w kolejnych realizacjach badania, 2021 Q1 – 2022 Q2



Źródło: Opracowanie własne na podstawie BAEL

Analiza wynagrodzeń w BAEL

Zmiany wynagrodzeń netto w kolejnych realizacjach badania, 2021 Q1 – 2022 Q2



Źródło: Opracowanie własne na podstawie BAEL

Uwaga 1.: Kolejność zmian wynagrodzenia malejąca według liczby wystąpień

Uwaga 2.: Na przykład wysokość słupka dla zmiany wynagrodzenia 500 oznacza, że 27,6% obserwacji zmian wynagrodzenia (z wyłączeniem zerowych zmian), stanowiły wartości 100, 200, 300 i 500

Propozycja procedury imputacji wynagrodzenia w BAEL

Imputacja w kwartale *t*:

1. Osobom, które nie podały wynagrodzenia, ale podały przedział wynagrodzenia imputowana jest wartość wynagrodzenia w oparciu o obserwowany rozkład wynagrodzenia w podanym przedziale.
2. Osobom, które nie podały wynagrodzenia, ale brały udział w badaniu we wcześniejszych kwartałach i były pytane o wynagrodzenie, przepisywane jest wynagrodzenie (podane przez respondenta lub zaimputowane)[1].
3. Osobom z poprzedniego kroku imputowana jest zmiana wynagrodzenia w oparciu o obserwowany rozkład zmian wynagrodzenia z próby, a następnie zaimputowana wartość dodawana jest do przepisanego w poprzednim kroku wynagrodzenia[1].
4. Pozostałym osobom, które nie podały wynagrodzenia i nie były pytane o wynagrodzenie w poprzednich kwartałach, imputowana jest wartość wynagrodzenia w oparciu o obserwowany rozkład wynagrodzenia z próby.

[1] dla $t = 1$ ten krok jest pomijany

Propozycja procedury imputacji wynagrodzenia w BAEL

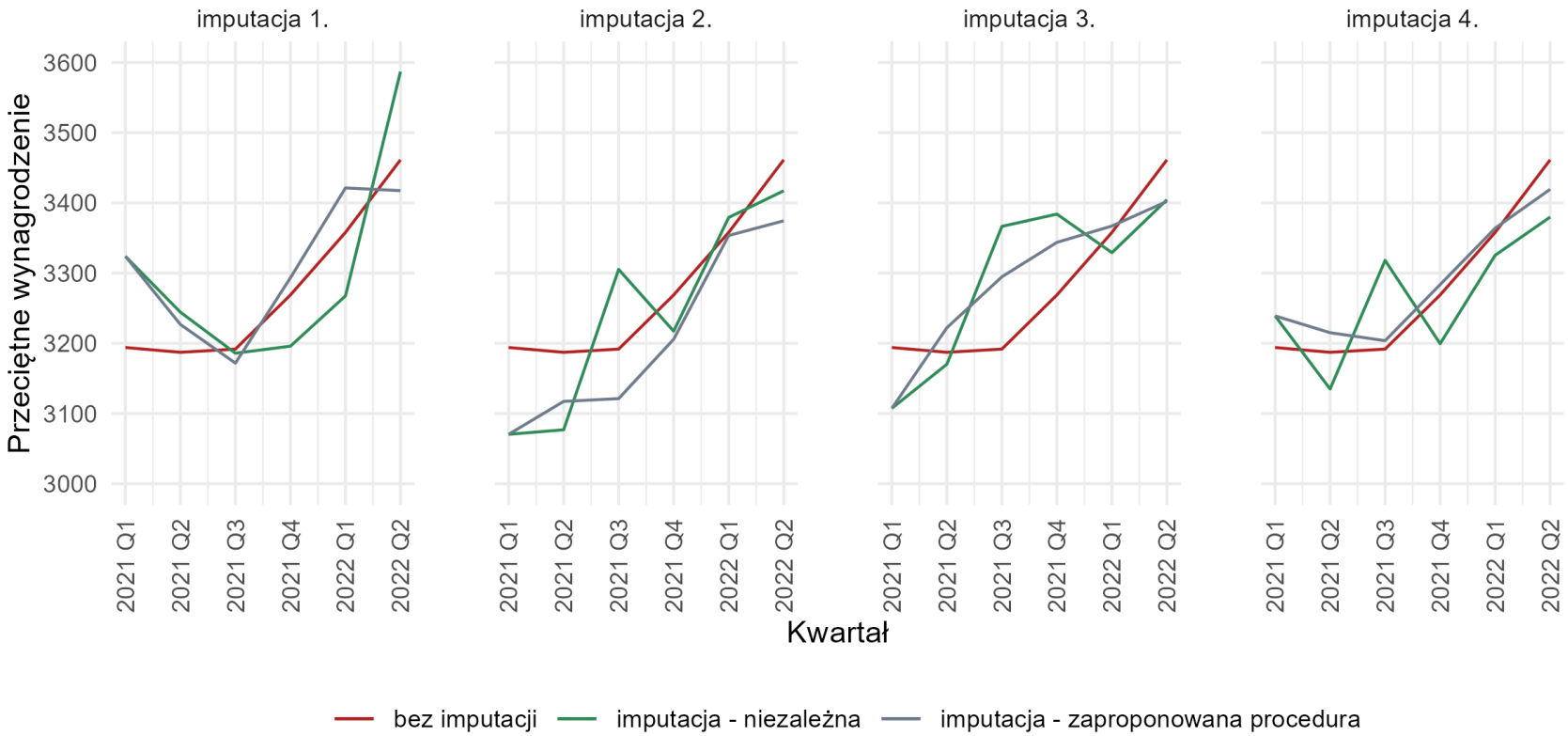
Metoda imputacji[1]

- PMM – Predictive mean matching (pakiet mice w języku R)
- zmienne pomocnicze: płeć, wiek, poziom wykształcenia, klasa miejscowości, województwo

[1] to nie jest ostateczny wybór metody imputacji, służy jedynie do egzemplifikacji zaproponowanej procedury.

Propozycja procedury imputacji wynagrodzenia w BAEL

Przeciętne wartości wynagrodzenia netto w próbie wyznaczone w oparciu o przykładowe realizacje imputacji



Źródło: Opracowanie własne na podstawie BAEL

Uwaga.: W imputacji niezależnej każdy kwartał był imputowany niezależnie od wartości z innych kwartałów

Dalsze kierunki badań

1. Symulacja w oparciu o dane syntetyczne.
2. Opracowanie metody bootstrap służącej do oceny precyzji oszacowań przy uwzględnieniu imputacji danych.
3. Przetestowanie różnych technik imputacji wraz z różnymi zestawami zmiennych pomocniczych i wybór optymalnej wersji imputacji.

Urząd Statystyczny
w Poznaniu

Dziękuję za uwagę