

SGH

Szkoła Główna
Handlowa
w Warszawie

Zastosowanie analizy sekwencji oraz dekompozycji APC do analizy danych retrospektywnych o pobycie imigrantów w Polsce

Paweł Strzelecki, SGH

4 lipca 2023, Konferencja Naukowa „Metodologia Badań Statystycznych” MET2023 Warszawa

Plan prezentacji

1. Motywacja

2. Analiza sekwencji

- Opis metody - analiza sekwencji
- Dane źródłowe
- Przykładowe wyniki modelowania

3. Hierarchiczny model logitowy z efektami kohortowymi

- Opis metody – APC na danych indywidualnych
- Dane źródłowe
- Przykładowe wyniki modelowania

4. Podsumowanie

Motywacja

1. Potrzeby analityczne:

- ⇒ trudności w ocenie (modelowaniu) przekazów pieniężnych
 - dynamiczne zmiany napływu migrantów
 - zróżnicowane strategie migracyjne

2. Dostępność danych:

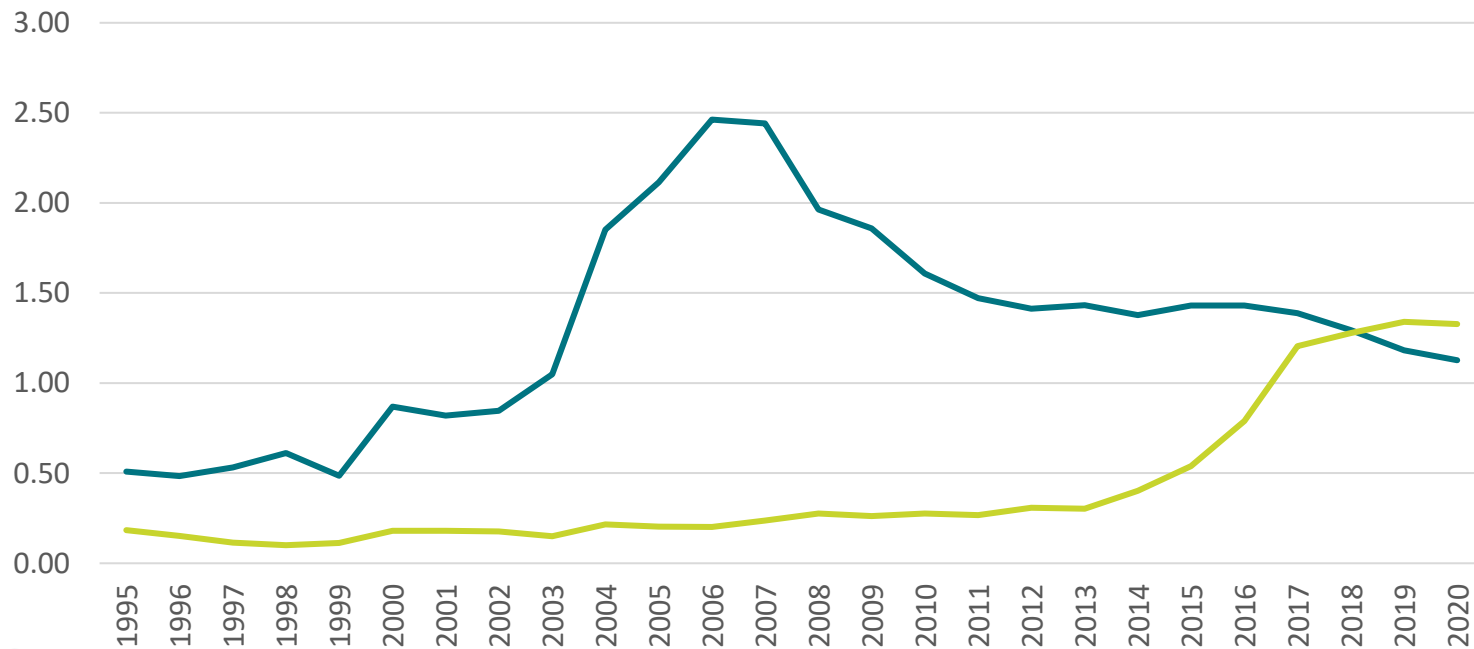
- ⇒ zmienne retrospektywne w badaniach ankietowych NBP

3. Cele naukowe:

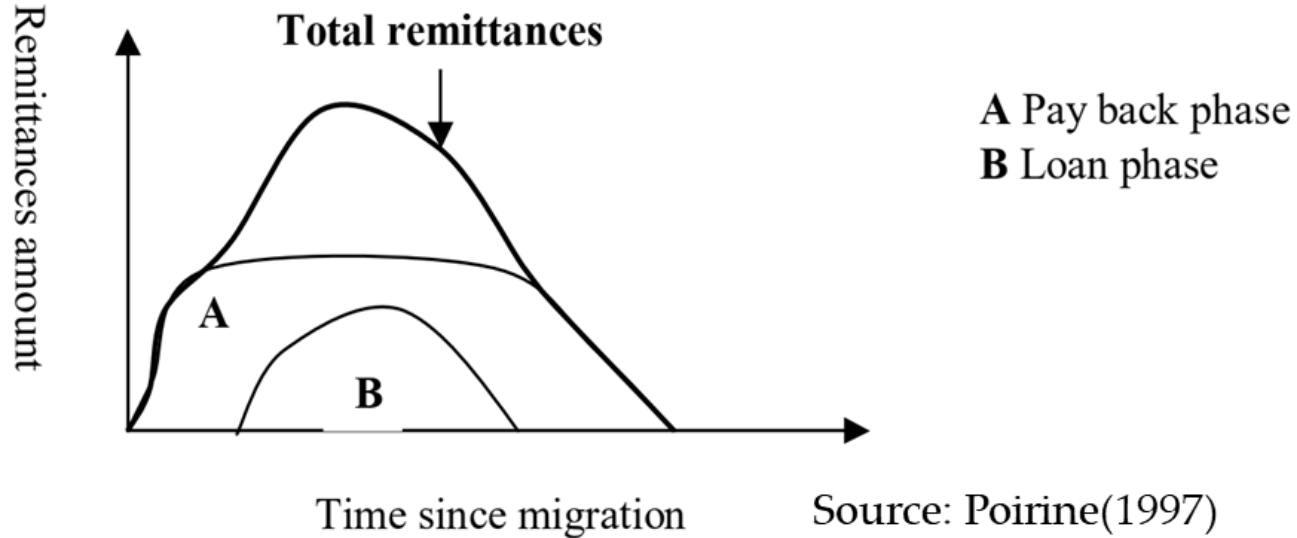
- ⇒ Transformacja od migracji krótkookresowej do migracji osiedleńczej
- ⇒ Weryfikacja „*remittances decay hypothesis*”

Remittances - znaczenie

Personal remittances as % of GDP in Poland



Remittances Decay Hypothesis



Wykorzystane typy danych retrospektywnych

- Badania ankietowe imigrantów w Polsce przeprowadzone przez Ośrodek Badań nad Migracjami UW dla NBP metodą *Respondent Driven Sampling* w latach 2018-2021

Ważne!

- Rozbudowana metryczka i pytania o przekazy pieniężne
- Szczegółowe pytania o przeszłość migracyjną w Polsce => **analiza sekwencji**
- Szczegółowe informacje dotyczące roku i miesiąca ostatniego przejazdu do Polski w kolejnych badaniach w 2018 (Wrocław), 2019 (Warszawa), 2020 (Wrocław) => **dekompozycja APC**

Przeszłość migracyjna – pytanie ankietowe

Karta odpowiedzi respondenta (przykład):

Rok	sty		lut		mar		kwi		maj		cze		lip		sie		wrze		paź		lis		gru	
1989																								
1990																								
1991																								
1992																								
.....																								
2015																								
2016																								
2017																								
2018								X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2019			X	X	X	X	X	X	X								X	X	X	X	X	X	X	X

SGH

Początek doświadczeń z migracjami do Polski

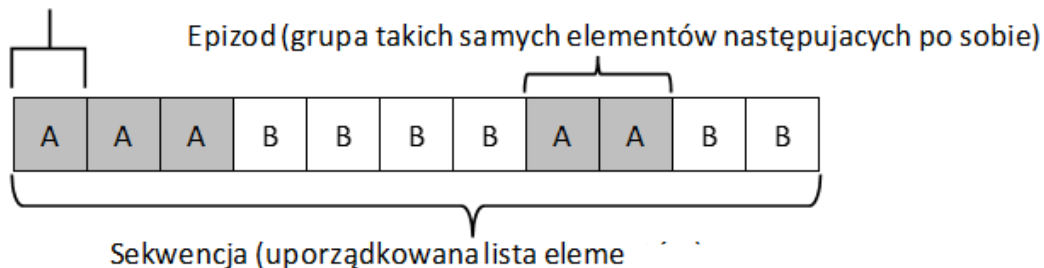
Ostatni pobyt migracyjny

www.sgh.waw.pl

Analiza sekwencji

Analiza sekwencji - metoda

Pojedyncze elementy (np: status migracyjny w każdym roku)



Osoba 1

A	A	A	B	B	B	B	A	A	B	B
---	---	---	---	---	---	---	---	---	---	---

Osoba 2

A	A	B	B	A	B	B	A	B	B	B
---	---	---	---	---	---	---	---	---	---	---

Odległości pomiędzy elementami

0	0	s	0	s	s	0	0	s	0	0
---	---	---	---	---	---	---	---	---	---	---

Źródło: Opracowanie własne na podstawie: (Brzinsky-Fay et al. 2006:1)

Wstępne grupowanie danych

Liczba osób z poszczególnymi kategoriami obecności w Polsce w zbiorze danych w latach 2014 - 2019

	2014	2015	2016	2017	2018	2019
Kat 0 - brak obecności w Polsce	1110	982	799	545	189	66
Kat 1 - obecność w Polsce 0 - 5 miesięcy	103	147	204	287	405	714
Kat 2 - obecność w Polsce 6-11 miesięcy	66	112	186	276	420	343
Kat 3 - obecność w Polsce cały rok (pełne 12 miesięcy)	40	78	130	211	305	196
RAZEM	1319	1319	1319	1319	1319	1319

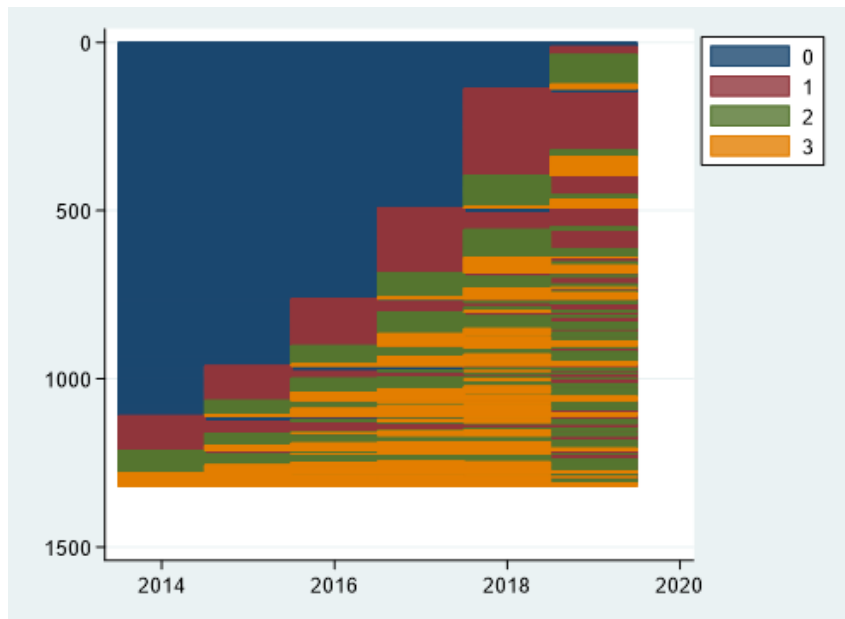
Wyniki – popularność sekwencji

Ranking sekwencji według częstości:	Sekwencja						Częstość występowania
	2014	2015	2016	2017	2018	2019	
1)	0	0	0	0	1	1	12.9%
2)	0	0	0	0	2	2	6.7%
3)	0	0	0	0	2	1	4.3%
4)	0	0	0	1	2	1	4.2%
5)	0	0	0	0	1	3	4.2%
6)	0	0	0	1	1	1	2.9%
7)	0	0	0	1	2	2	1.9%
8)	0	0	0	0	0	1	1.8%
9)	0	0	0	1	2	2	1.7%
10)	0	0	0	0	2	3	1.6%
RAZEM:							42.1%

SGH

0-brak obecności w Polsce
1- obecność w PL 1-5 m-cy
2- obecność 6-11 m-cy
3- obecność pełne 12 m-cy

Wyniki – graficzna prezentacja sekwencji

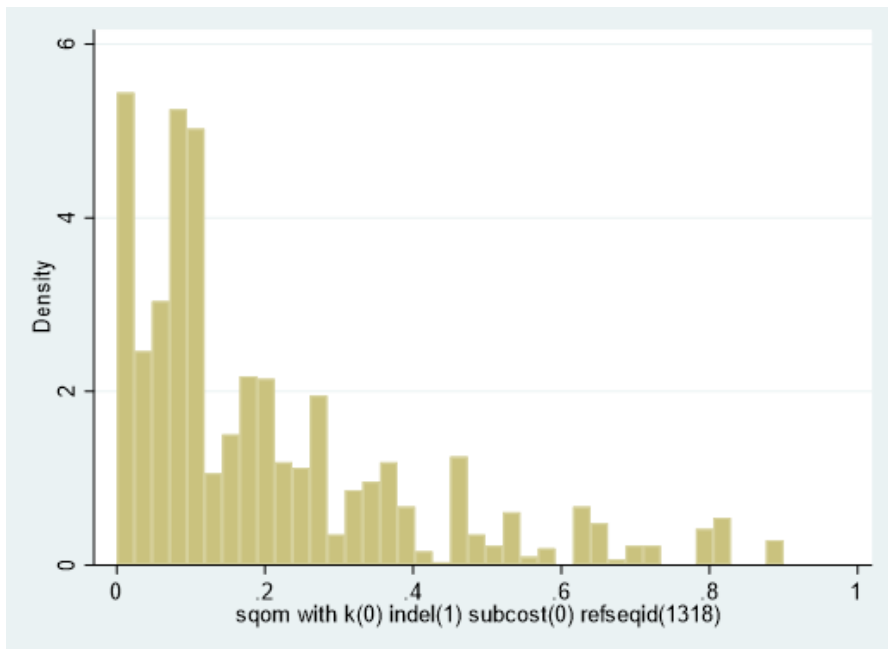


Odległości pomiędzy kategoriami - założenia

		Drugi porównywany element			
		0	1	2	3
Pierwszy porównywany element	0	0.0	0.3	0.5	1.0
	1		0.0	0.2	0.7
	2			0.0	0.5
	3				0.0

Wyniki – odległość od sekwencji referencyjnej

Sekwencja referencyjna: najczęstsza sekwencja w zbiorze danych [0 0 0 0 1 1]

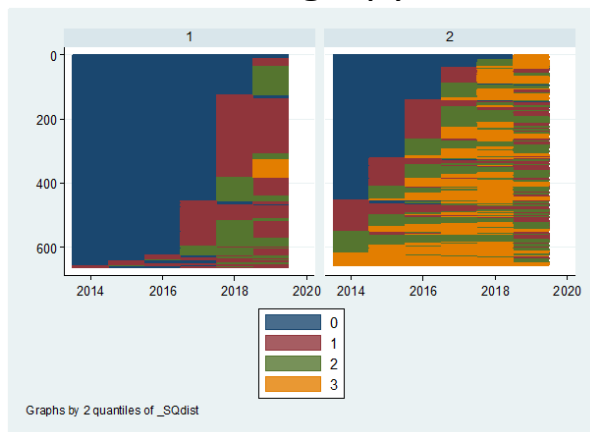


Odległości pomiędzy
kategoriami - założenia

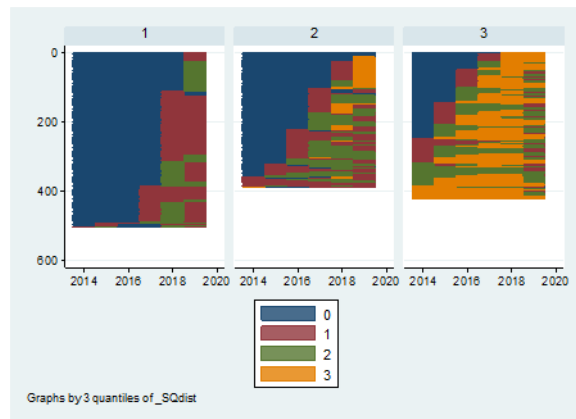
		Drugi porównywany element			
		0	1	2	3
Pierwszy porównywany element	0	0.0	0.3	0.5	1.0
	1		0.0	0.2	0.7
	2			0.0	0.5
	3				0.0

Analiza sekwencji - segmentacja

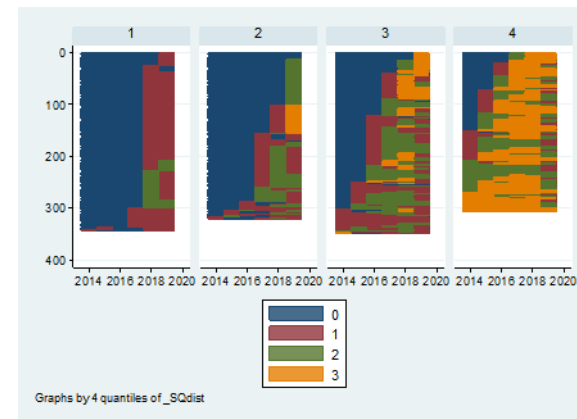
Dwie grupy



Trzy grupy



Cztery grupy

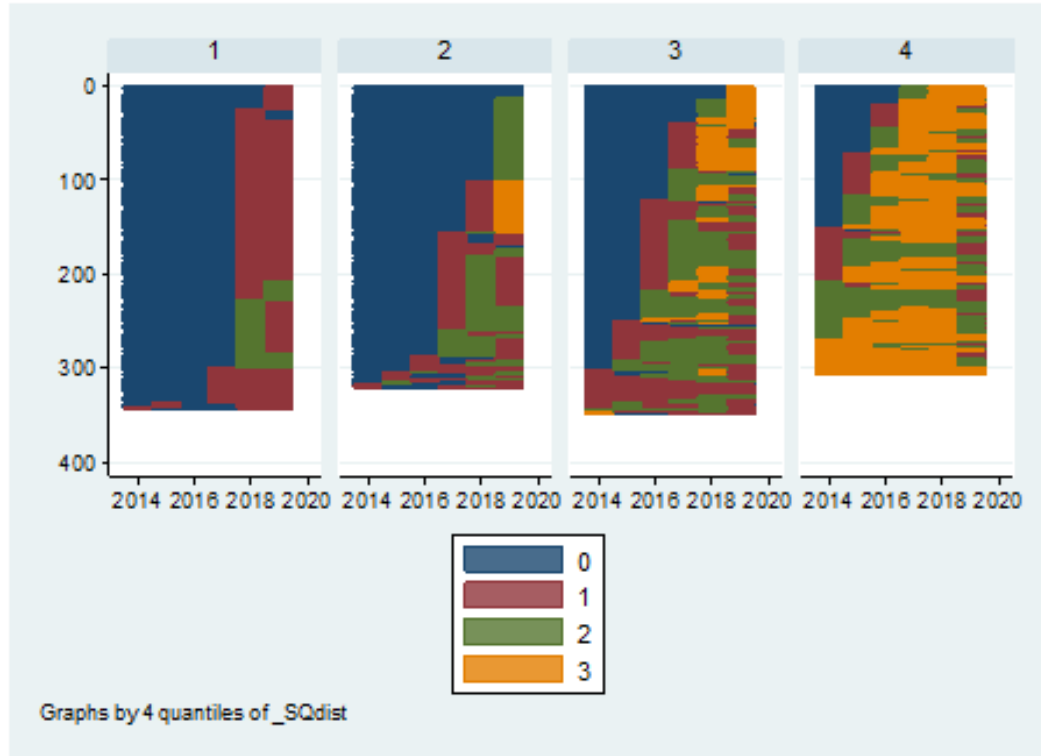


Grupowanie według sekwencji kategorii pozostawania w Polsce

1: Niedawni przyjezdni pozostający na krótko

2: Niedawni, ale „stabilni” przyjezdni lub z długim okresem nieobecności w Polsce

SGH



3: Pobyt krótki, ale regularny lub długi, ale nieregularny

4: Stabilny pobyt min 3 lata.

Strategia migracyjna a przekazywanie pieniędzy za granicę

Wyniki modelu logitowego:

- Zmienna objaśniana:
0/1 przesyłanie przekazów
- Modele:
- bazowy, informacja w postaci odległości, informacja w postaci grupowania
- raportowane odds ratios

SGH

Zmienne objaśniające	Model bazowy	Model z odległościami	Model z grupowaniem
	0.776** (0.0951)	0.763** (0.0953)	0.764** (0.0954)
kobieta (ref. Mężczyzna)			
wiek	1.060** (0.0313)	1.075** (0.0321)	1.074** (0.0317)
wiek2	1.000 (0.000384)	0.999 (0.000385)	0.999 (0.000379)
wykształcenie średnie (ref. zawodowe i mniej)	1.352 (0.354)	1.402 (0.373)	1.371 (0.366)
wykształcenie wyższe (ref. zawodowe i mniej)	0.706 (0.184)	0.779 (0.207)	0.751 (0.200)
Cluster 2 (ref. Cluster 1)			0.520*** (0.0931)
Cluster 3 (ref. Cluster 1)			0.496*** (0.0884)
Cluster 4 (ref. Cluster 1)			0.302*** (0.0544)
SQ - odległość od sekwencji najpopularniejszej		0.0157*** (0.0142)	
SQ2 - kwadrat odległości od sekwencji najpopularniejszej		24.84*** (29.93)	
Stała	0.521 (0.300)	0.646 (0.377)	0.720 (0.421)
Liczba obserwacji	1,316	1,316	1,316

Informacja o istotności statystycznej:
*** p<0.01, ** p<0.05, * p<0.1

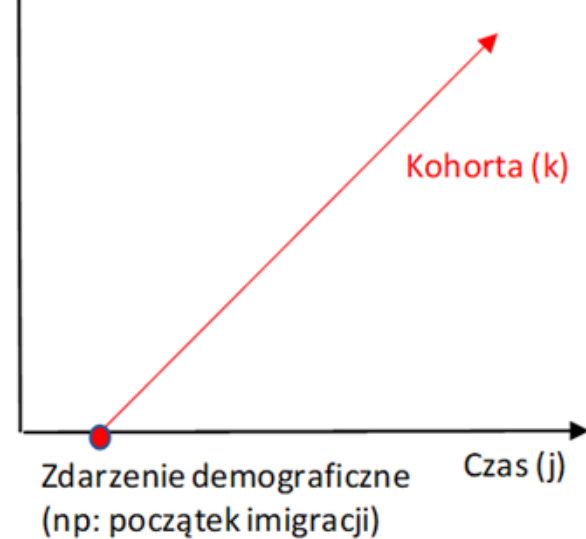
Źródło: Opracowanie własne

Analiza APC na danych indywidualnych

Hierarchiczny logit z efektami kohortowymi - metoda

Kohorta migracyjna

Wiek (i) „wiek” – długość pozostawania w Polsce od momentu migracji



„kohorta” – osoby, które przyjechały do Polski w tym samym czasie

Uproszczona wersja modelu

Stosujemy podejście oparte na Yang & Land (2008): hierarchiczny model logitowy



$$\ln\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \beta_0 + \beta_1 \text{Duration}_{ijk} + u_j + u_k$$

- π_{ijk} jest prawdopodobieństwem przesyłania pieniędzy dla osoby i , w okresie j i kohorcie migracyjnej k .
- u_j efekty stałe (okres/miejsce) oraz u_k są efektami losowymi w odpowiednim okresie lub kohorcie migracyjnej

Zastosowanie do danych

Zmienna objaśniana:

- wysyłanie przekazów (0/1)

Główne zmienne objaśniające:

- „Duration” – okres pozostawania imigrantów w Polsce
- Zmienne zerojedynkowe dla czasu/miejsca obserwacji w Polsce
- Kohorta migracyjna (2014-2021)

Zmienne kontrolne:

- płeć (kobieta=1)
- wiek (ciągła)
- wykształcenie (3 poziomy)

Wyniki modelowania

remit	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
female01	.6927721	.0644691	-3.94	0.000	.5772686	.8313864
age	1.07563	.0258759	3.03	0.002	1.026091	1.12756
age2	.9993493	.0003132	-2.08	0.038	.9987357	.9999634
edu3						
1	1.300314	.2640717	1.29	0.196	.8733403	1.936033
2	1.743395	.1680578	5.77	0.000	1.443254	2.105954
_Iyear_2019	2.004776	.2522554	5.53	0.000	1.566612	2.565489
Iyear_2020	1.380662	.2108118	2.11	0.035	1.023572	1.862329
duration	.9843523	.0036318	-4.27	0.000	.9772598	.9914963
_cons	.2756699	.1279345	-2.78	0.005	.1110092	.6845731

Note: **_cons** estimates baseline odds (conditional on zero random effects).

Random-effects parameters	Estimate	Std. err.	[95% conf. interval]	
migcohort: Identity				
var(_cons)	.0094749	.0154929	.0003843	.233573

Podsumowanie

- Metody statystyczne oparte o **indywidualne dane retrospektywne** pozwalają na spojrzenie na różne zjawiska z perspektywy doświadczeń migracyjnych (**trajektorii migracji**).
- **Analiza 1:** częstości przesyłania środków pieniężnych (**remittances**) według doświadczeń migracyjnych **zależą od klasyfikacji migrantów** opartej na czasie pozostawania w Polsce.
- **Analiza 2:** Modelowanie wielopoziomowe z efektami losowymi pozwala na uwzględnienie potencjalnych „**efektów kohortowych**” i „efektów czasowych” w objaśnianiu częstości przesyłania środków.