

Model grawitacyjny migracji

RAPORT ZALICZENIOWY – METODY EKONOMETRYCZNE
ALEKSANDRA KACPRZYK 114125

Spis treści

1.	Wprowadzenie.....	2
2.	Opis danych.....	4
2.1	Zmienna objaśniana.....	4
2.2	Zmienne objaśniające	5
3.	Model	11
3.1	Model 1	11
3.2	Model 2	12
3.3	Model 3	14
4.	Analiza ekonometryczna	16
4.1	Oszacowania parametrów modelu 3 MNK i ich interpretacja	16
4.2	Test RESET	16
4.3	Współliniowość zmiennych objaśniających.....	16
4.4	Normalność rozkładu składnika losowego.....	17
4.5	Heteroskedastyczność	18
4.6	Metody uwzględniające niesferyczność składnika losowego	19
5.	Podsumowanie	21
6.	Bibliografia.....	22

1. Wprowadzenie

Według wyliczeń Międzynarodowej Organizacji ds. Migracji (IOM) dla roku 2020 było około 281 milionów migrantów, co jest równoznaczne z tym, że co trzydziesta osoba jest migrantem ('World Migration Report 2024', 2020). Mimo wrażenia, że to nie jest duża liczba, warto zauważyć, że na początku XX wieku to było około 174 milionów migrantów ('World Migration Report 2024', 2020). Zjawisko to charakteryzują się trendem rosnącym, co może być powiązane z zwiększaniem intensywności zjawiska globalizacji. Jest to obszar warty przeanalizowania, nie tylko z powodu jego trendu, ale również przyczyn społeczno-ekonomicznych, na przykład, ten temat jest niezbędny w usprawnieniu polityki wobec migrantów.

W niniejszym raporcie będzie poddany rozważaniom model grawitacyjny dla migracji. Jest to rodzaj modelu, który od wielu lat był rozpoznawany za wysoką skuteczność w wyjaśnianiu przepływów (Bergstrand, 1985). Jest on oparty na pewnym założeniu, że grawitacja jest wszechobecna i opiera się na uniwersalnym prawie powszechnego ciążenia. Poniżej jest przedstawiony wzór, który w fizyce służy do obliczenia siły grawitacji:

$$F = G \frac{m_1 m_2}{r^2}$$

Jednak warto wziąć pod uwagę, że ludzie i ich decyzje, które wpływają na to jak kształtuje się świat w sensie polityczno-społeczno-ekonomicznym, są zmienne i zależne od większej ilości czynników niż prawo powszechnego ciążenia (Crymble, 2019). Zatem, w tym raporcie będziemy rozważać kształtowanie się teorii bazującej na założeniu, że siłę migracji można wytłumaczyć używając czynników społeczno-ekonomicznych.

W swojej podstawowej formie model grawitacyjny opisuje logarytm przepływu z kraju pochodzenia (miejsca początkowego) do kraju docelowego (miejsca docelowego) za pomocą logarytmów wartości społeczno-ekonomicznych. Co więcej, charakteryzuje się on wysokim wyjaśnieniem determinant modelu (Anderson, 2011). Jednak nie powinno się na jego podstawie przeprowadzać prognoz, co jest związane z częstym brakiem podstaw teoretycznych (Bergstrand, 1985).

Pierwszym opisanym modelem grawitacyjnym był model dotyczący migracji z jednego miejsca do innego. To było związane z przekonaniem „*Im większa jest liczba (ludzi) zebranych na danym obszarze, tym większa powstaje tam siła przyciągania...*” (Chojnicki, 1966). Na początku zauważono, że wielkość ruchu migracyjnych zależy od odległości między miejscem

początkowym a docelowym. Co więcej, w tym modelu też musi istnieć, tak zwana siła oddziaływania, co jest odpowiednikiem siły tych jednostek terytorialnych (Chojnicki, 1966). Mierzono to za pomocą wielkości popytu artykułu. Z czasem zaczęto używać liczby ludności w tych miejscach. Model, więc przybrał formę równania: $M_{1,2} = \frac{L_1 L_2}{D}$, gdzie M to przepływy migracji między krajem pochodzenia (1) a krajem docelowym (2), L_1 – ludność w kraju 1, L_2 – ludność w kraju 2, D – odległość między krajem 1 a krajem 2 (Chojnicki, 1966). Jednak, należy załączyć do tego modelu siłę atrakcyjności miejsca docelowego, gdyż musi być powód podjęcia decyzji o migracji tych osób. Zatem model został zmodyfikowany w celu dołączenia do niego motywacji stojących za tym zachowaniem (Bergstrand, 1985). Jednak z powodów technicznych, w niniejszym raporcie zdecydowano się skoncentrować się na motywacjach finansowych, czyli na migracji zarobkowej. Zatem w zmodyfikowanym modelu, niektórzy uwzględniają pensje.

Celem niniejszego raportu jest przedstawienie i przeanalizowanie uzyskanych zależności między przepływami migracji między krajem pochodzenia a krajem docelowym, a populacją tych obu państw i ich GNI *per capita* (Gross National Income *per capita*, po polsku: dochód narodowy *per capita*), i fizycznym dystansem między nimi. Zatem, odpowiada on na pytanie badawcze: „Czy można wyjaśnić przepływy migracji przy pomocy populacji, dochodu narodowego obu krajów i dystansu między nimi? Jaka jest zależność między nimi?”.

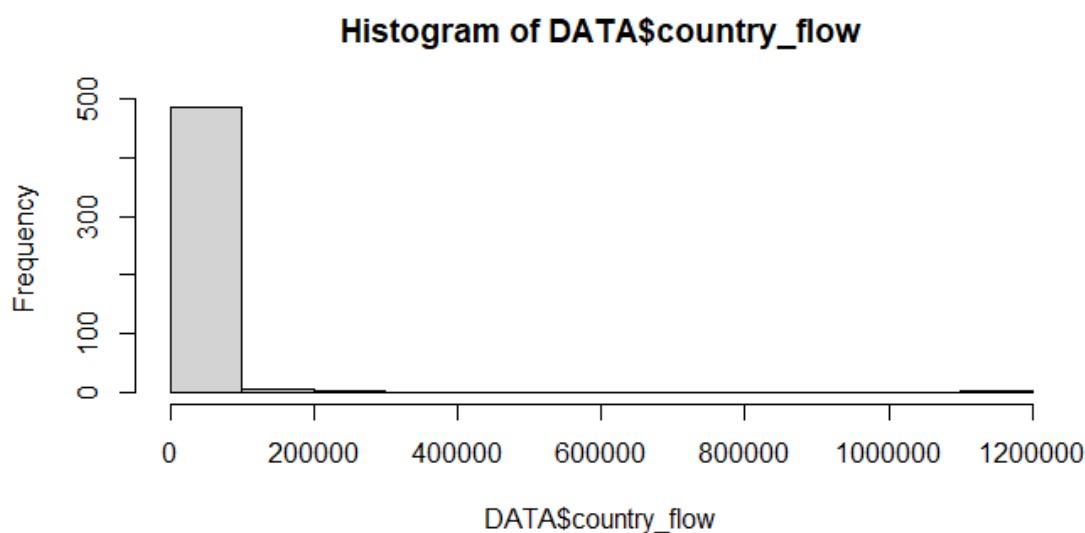
2. Opis danych

Dane użyte w modelu pochodzą z bazy Banku Światowego, CEPI (francuska instytucja zajmująca się badaniem i ekspertyzą ekonomii światowej), badania dotyczące globalnych przepływów ludności publikacji Abela and Sandera (2014) pod tytułem: „*Quantifying Global International Migration Flows*”. Zestaw danych po agregacji pojedynczych baz danych i oczyszczeniu niezbędnym do późniejszej analizy zawiera 494 obserwacje. Zdecydowano się usunąć obserwacje dla których zmienna objaśniana przyjmuje wartości zerowe. Biorąc pod uwagę charakterystykę zmiennej objaśnianej, dane pochodzą z przekroju lat 2005 – 2010. Jednak by zachować jednolitość zmiennych zdecydowano się na użycie pewnej modyfikacji, która jest poniżej opisana.

2.1 Zmienna objaśniana

Przepływy migracji (*country_flow*) jest to zmienna, która mierzy w milionach zmianę w liczbie migrantów dla pięcio-letniego okresu, od 2005 do 2010 z podziałem kraj-do-kraju, czyli z informacją z jakiego kraju pochodzenia do jakiego kraju docelowego nastąpił przepływ. Pochodzą one z publikacji Abela and Sandera (2014) pod tytułem: „*Quantifying Global International Migration Flows*”. Są one zabrane dla 196 krajów.

WYKRES 1. HISTOGRAM ZMIENNEJ OBJAŚNIANEJ *COUNTRY_FLOW*

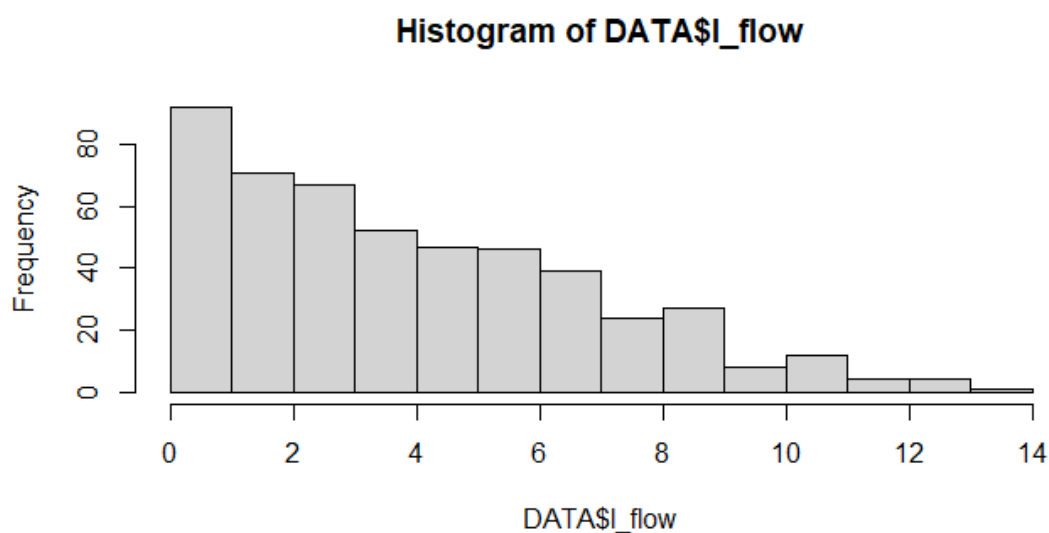


ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Wykres 1 przedstawia histogram zmiennej objaśnianej, czyli przepływów migracji między krajami. Rozkład zmiennej charakteryzuje się lewo-skością. Większość obserwacji

znajduje się w przedziale od 1 do 100 000. Biorąc pod uwagę rozkład tej zmiennej zdecydowano się na utworzenie logarytmu tej zmiennej i nazwano nową zmienną l_flow . Cechuje się również rozkładem lewo-stronnym, jednak jest większe rozproszenie obserwacji.

WYKRES 2. HISTOGRAM ZMIENNEJ OBJAŚNIANEJ L_FLOW



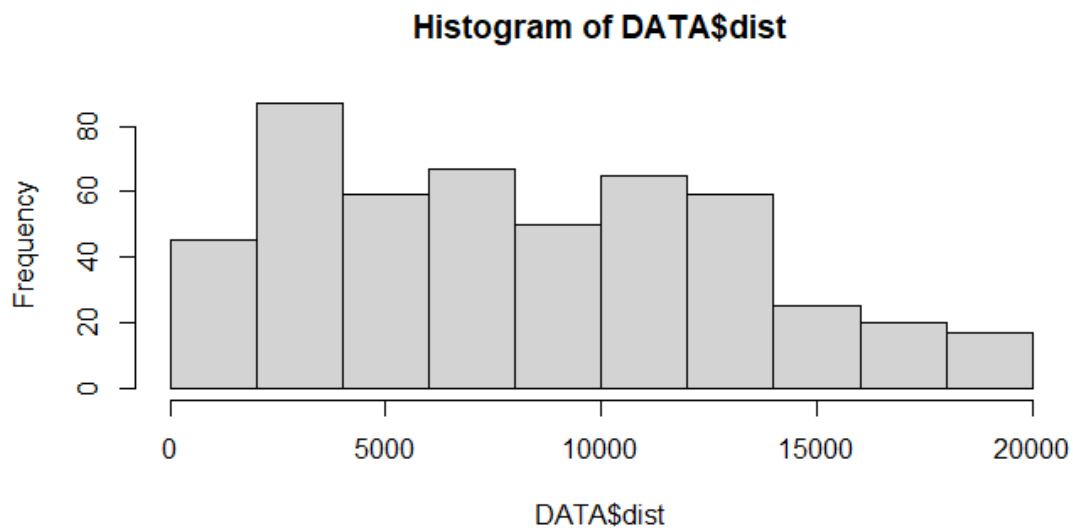
ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

2.2 Zmienne objaśniające

Poniżej opisaną każdą zmienną objaśniającą modelu:

- Dystans między krajem pochodzenia a krajem docelowym (*dist*) jest zmienną ciągłą, która opisuje odległość w kilometrach między miastami o największej liczbie ludności. Przyjmuje ona tylko wartości dodatnie. Poniższy wykres 3 przedstawia histogram tej zmiennej. Rozkład cechuje się lewo-stronnością.

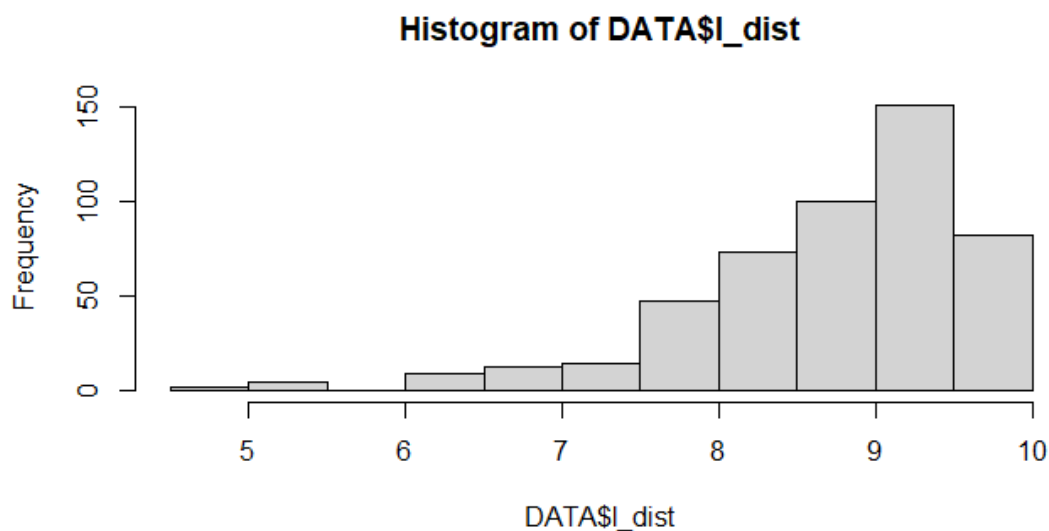
WYKRES 3. HISTOGRAM ZMIENNEJ *DIST*



ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Według literatury logarytmu zmiennej mierzącej dystans, dlatego zdecydowano się go utworzyć i nazwano go *l_dist*. Zgodnie z histogramem przedstawionym na wykresie 4 cechuje się on prawo-skośnością. Większość obserwacji przypada na przedział od 9 do 9,5.

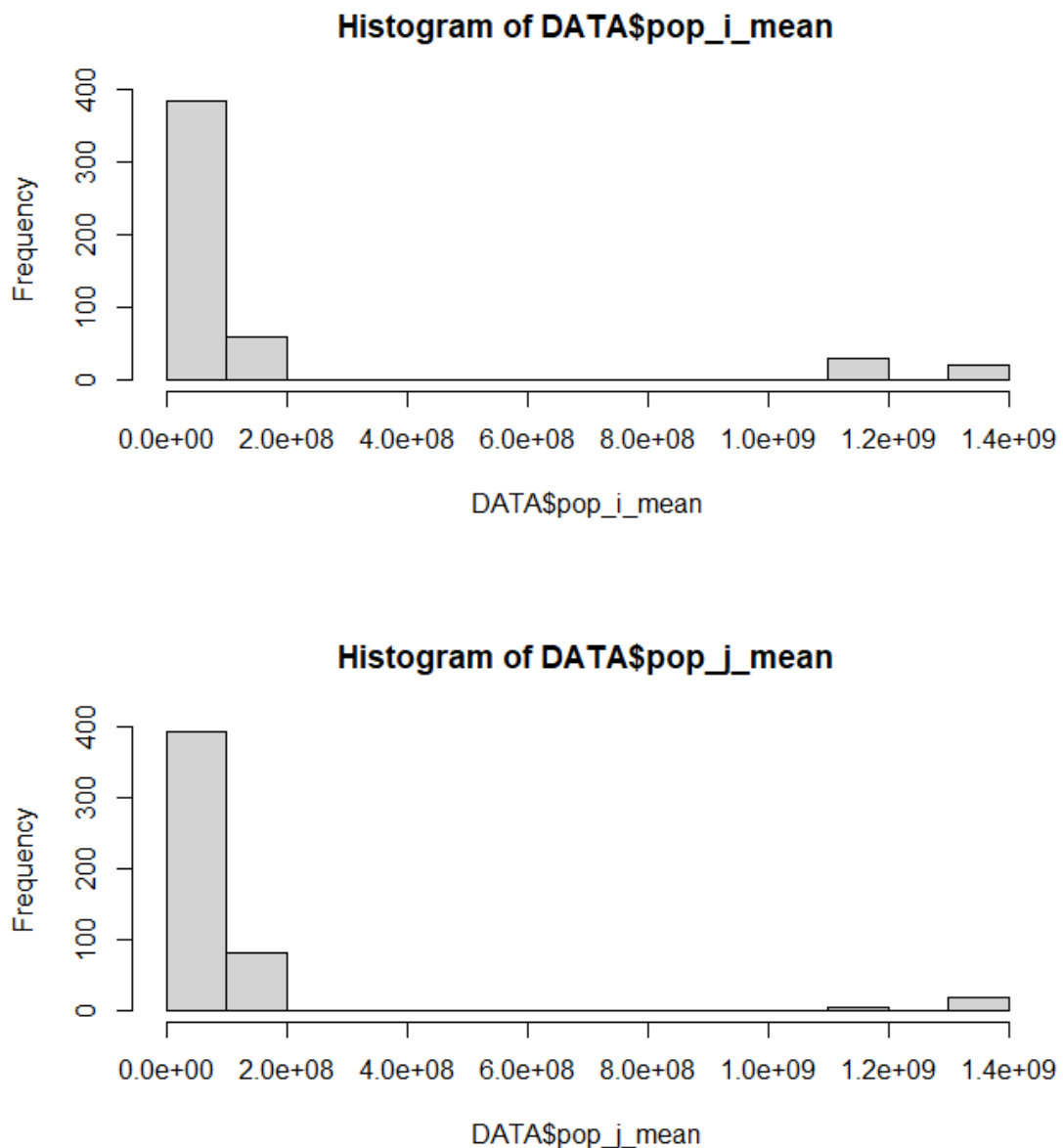
WYKRES 4. HISTOGRAM ZMIENNEJ *L_DIST*



ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

- Następnymi zmiennymi są: populacja kraju pochodzenia (*pop_i_mean*) i populacja kraju docelowego (*pop_j_mean*). Pierwsza z nich opisuje średnią populację w kraju pochodzenia z lat 2005-2010. Był to zabieg zastosowany w celu dopasowania populacji do zmiennej objaśnianej. Analogicznie jest dla drugiej zmiennej. Poniżej przedstawiono histogramy dla tych dwóch zmiennych. Oba charakteryzują się lewo-skośnością i większość obserwacji znajduje się w pierwszym przedziale.

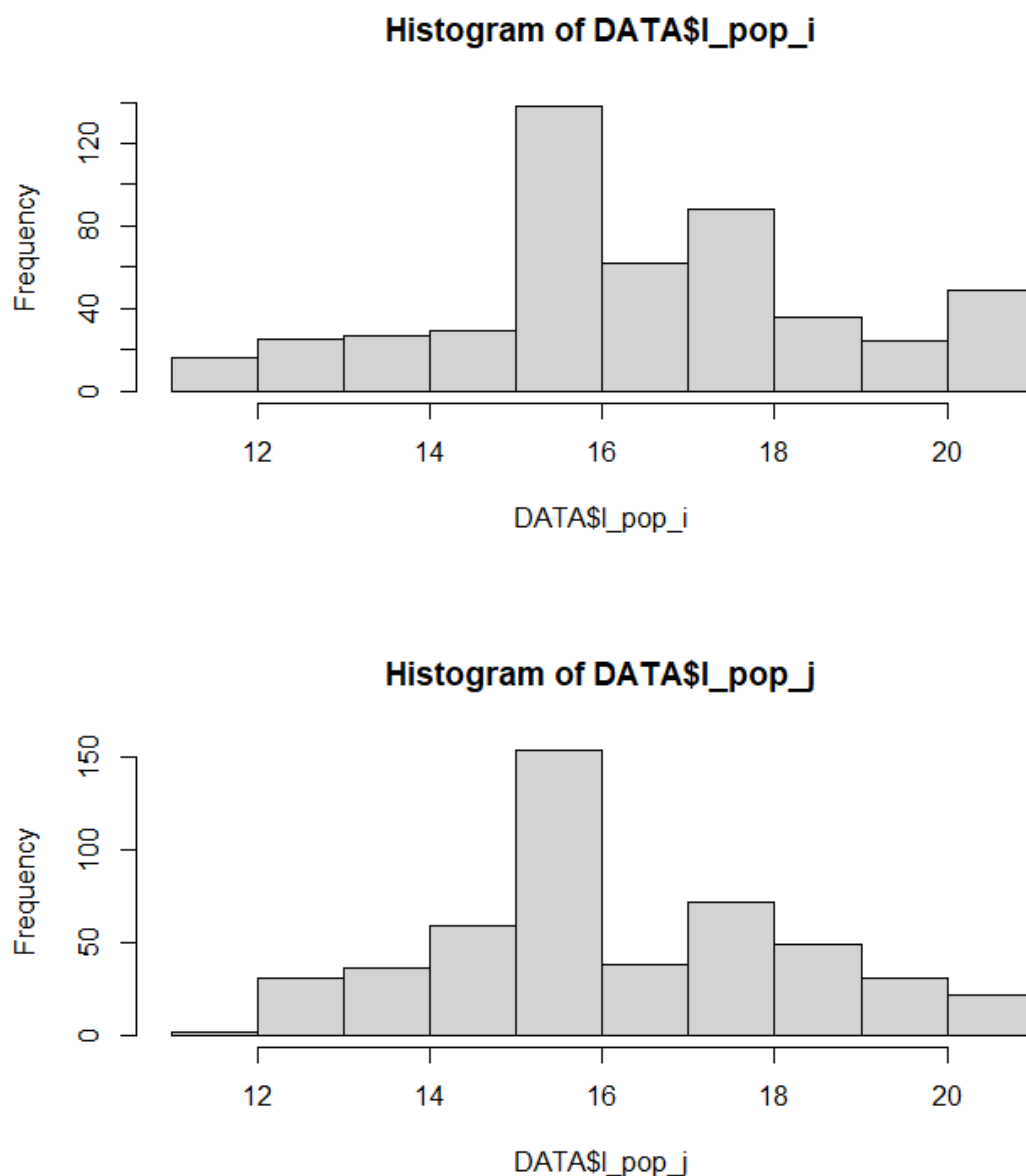
WYKRES 5. HISTOGRAMY ZMIENNYCH *POP_I_MEAN* I *POP_J_MEAN*



ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Zgodnie z literaturą utworzono logarytmy tych zmiennych nazwane kolejno: l_pop_i i l_pop_j . Histogramy nazwane *wykres 6* przypominają rozkład normalny.

WYKRES 6. HISTOGRAMY ZMIENNYCH L_POP_I I L_POP_J

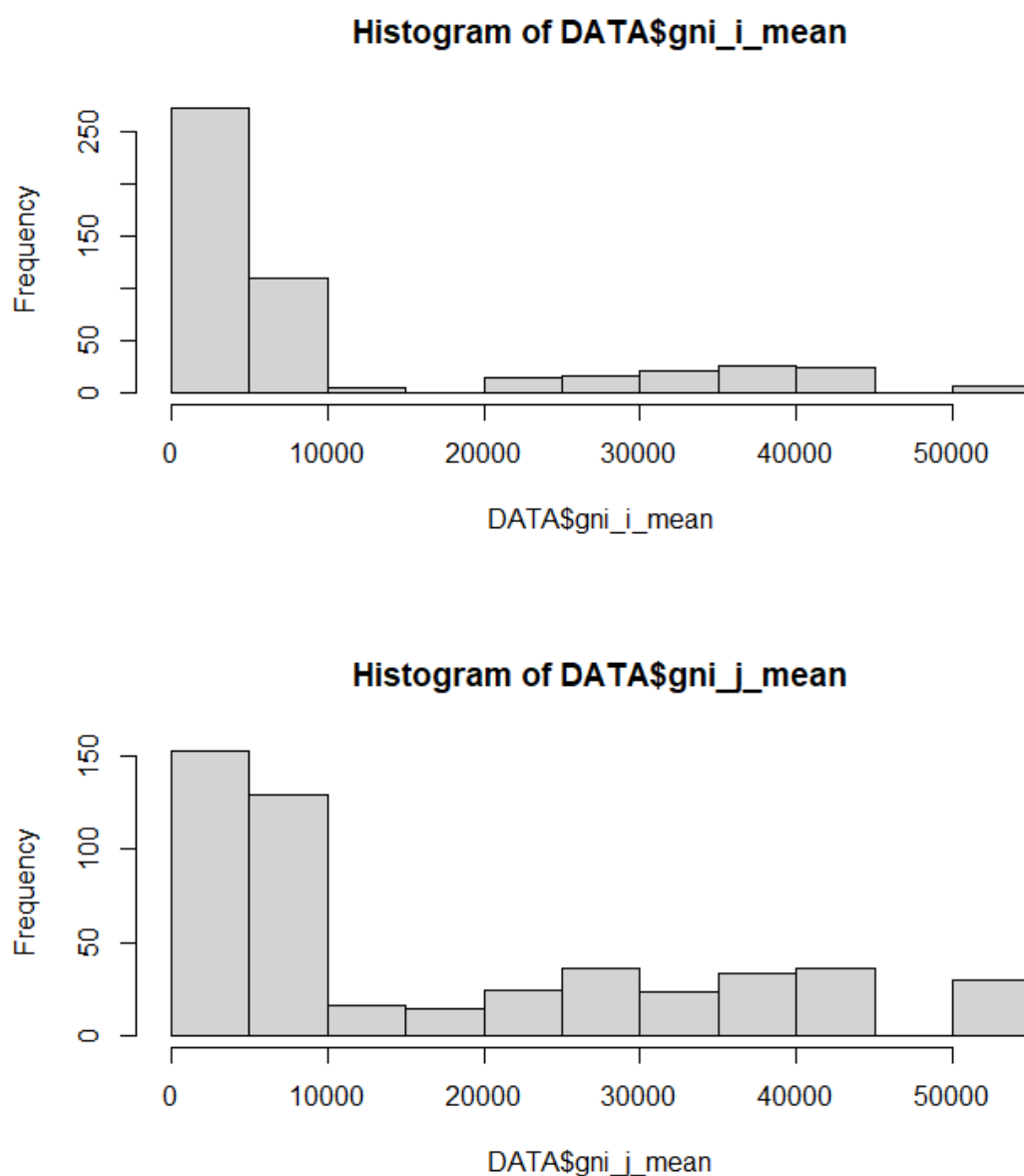


ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

- Ostatnimi zmiennymi są: dochód narodowy kraju pochodzenia (gni_i_mean) i dochód narodowy kraju docelowego (gni_j_mean). Pierwsza z nich zawiera średnią dochodu narodowego per capita z lat 2005-2010 wyrażonego w obecnych dolarach USD. Podobnie jest dla drugiej zmiennej. Zdecydowano się na użycie dochodu narodowego zamiast płac z powodów technicznych. Jednak, wybrano dochód narodowy zamiast

produktu krajowego brutto, ponieważ ten pierwszy obejmuje cały uzyskany dochódów otrzymany od obywateli czy firmy niezależnie od ich lokalizacji.

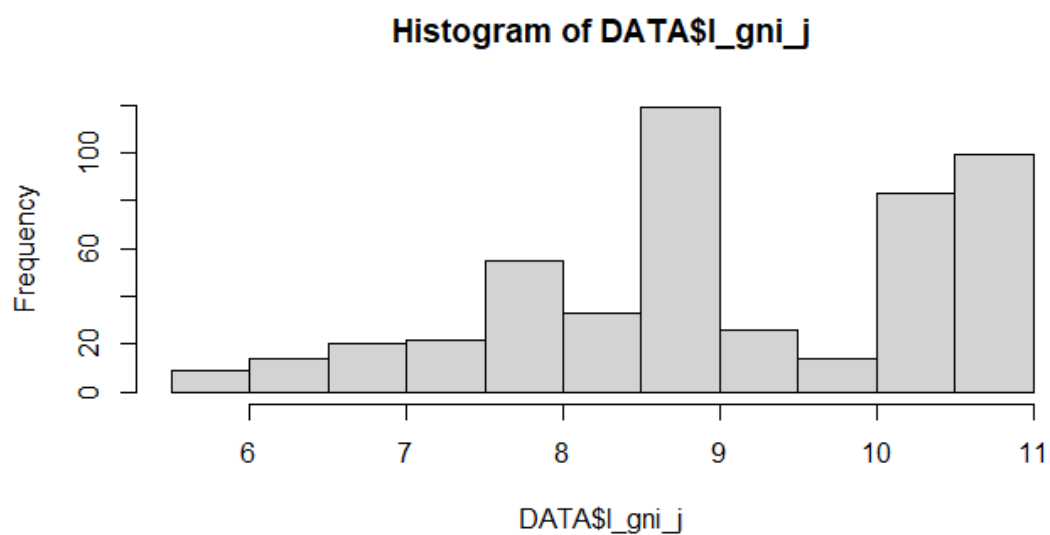
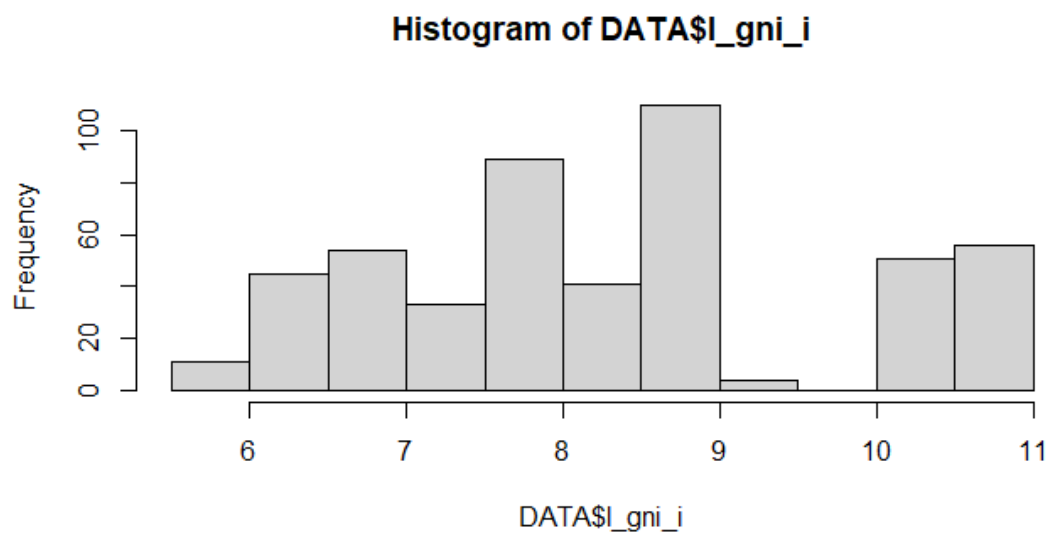
WYKRES 7. HISTOGRAMY ZMIENNYCH GNI_I_MEAN I GNI_J_MEAN



ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Histogramy w wykresie 6 są lewo-stronne i większość ich obserwacji jest w przedziale od 0 do 500. Biorąc pod uwagę literaturę utworzono logarytmy z tych zmiennych, odpowiednio: l_gni_i i l_gni_j . Ich histogramy charakteryzują się prawo-stronnością. Dla obu najwięcej obserwacji przypada na przedział od 8,5 do 9.

WYKRES 8. HISTOGRAMY ZMIENNYCH L_GNI_I I L_GNI_J



ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

3. Model

3.1 Model 1

Na początku zdecydowano się na stworzenie modelu naiwnego. Użyto do niego wszystkich pierwotnych zmiennych, zgodnie z poniższym wzorem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5,$$

gdzie oznaczenia kolejno znaczą: y – country_flow, x_1 – pop_i_mean, x_2 – pop_j_mean, x_3 – dist, x_4 – gni_i_mean, x_5 – gni_j_mean. Poniżej przedstawiono wyniki oszacowań modelu.

TABELA 1. OSZACOWANIA MODELU 1

```
lm(formula = country_flow ~ pop_i_mean + pop_j_mean + dist +
    gni_i_mean + gni_j_mean, data = DATA)

Residuals:
    Min       1Q   Median       3Q      Max
-63725   -8807   -1764    3873  1086916

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.800e+03  5.808e+03   0.654  0.51331
pop_i_mean    3.747e-05  6.937e-06   5.402 1.03e-07 ***
pop_j_mean    4.670e-06  9.539e-06   0.490  0.62469
dist         -1.331e+00  5.135e-01  -2.592  0.00983 **
gni_i_mean     6.964e-03  1.838e-01   0.038  0.96980
gni_j_mean     4.406e-01  1.565e-01   2.816  0.00506 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 54880 on 488 degrees of freedom
Multiple R-squared:  0.0775,    Adjusted R-squared:  0.06805
F-statistic: 8.2 on 5 and 488 DF,  p-value: 1.894e-07
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Według oszacowanie nie wszystkie zmienne są statystycznie istotne i na przepływy migracji mają wpływ tylko zmienne opisujące: populację kraju pochodzenia, dystans i dochód narodowy dla kraju docelowego. Co się nie zgadza z teorią, gdyż populacje obu krajów powinny być statystycznie istotne w modelu.

Następnie przeprowadzono test poprawności specyfikacji modelu (test RESET) i test normalności składnika losowego Jarque-Bery. Poniżej umieszczono wyniki tych testów.

GRAFIKA 1. TEST RESET

RESET test

```
data: model1
RESET = 43.998, df1 = 2, df2 = 486, p-value < 2.2e-16
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

GRAFIKA 2. TEST JARQUE-BERA

Jarque Bera Test

```
data: ehat1
x-squared = 2108203, df = 2, p-value < 2.2e-16
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Dla obu testów hipoteza zerowa została odrzucona przy poziomie istotności równym 0,01. Co za tym idzie, składnik losowy modelu nie ma rozkładu normalnego według testu Jarque-Berego oraz nie ma poprawnej specyfikacji modelu według testu RESET. Zatem zdecydowano się zmodyfikować ten model.

3.2 Model 2

Ze względu na niepoprawną specyfikację i brak normalności składnika losowego, dokonano pewnych modyfikacji w modelu. Obecnie jest zgodny z poniższym wzorem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5,$$

gdzie oznaczenia kolejno znaczą: y – country_flow, x_1 – l_pop_i, x_2 – l_pop_j, x_3 – dist, x_4 – gni_i_mean, x_5 – gni_j_mean. Zmiana ta została podyktowana rokładem tych zmiennych oraz literaturą, gdzie użyto logarytmów dla populacji obu krajów. Poniżej przedstawiono wyniki oszacowań modelu 2.

TABELA 2. OSZACOWANIA MODELU 2

```
lm(formula = l_flow ~ l_pop_i + l_pop_j + dist + gni_i_mean +
    gni_j_mean, data = DATA)

Residuals:
    Min       1Q   Median       3Q      Max
-5.6286 -1.8387 -0.0763  1.6242  6.9588

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.185e+01  1.339e+00  -8.849  < 2e-16 ***
l_pop_i      5.107e-01  4.948e-02  10.322  < 2e-16 ***
l_pop_j      4.688e-01  5.672e-02   8.266  1.32e-15 ***
dist        -2.052e-04  2.334e-05  -8.792  < 2e-16 ***
gni_i_mean   1.566e-05  8.123e-06   1.928   0.0545 .
gni_j_mean   7.141e-05  7.228e-06   9.880  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.472 on 488 degrees of freedom
Multiple R-squared:  0.3315,    Adjusted R-squared:  0.3246
F-statistic: 48.39 on 5 and 488 DF,  p-value: < 2.2e-16
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

W celu sprawdzenie poprawności modelu 2, przeprowadzono kolejno test RESET i test Jarque-Bera. Poniżej przedstawiono ich wyniki.

GRAFIKA 3. TEST RESET

```
RESET test

data:  model2
RESET = 3.4458, df1 = 2, df2 = 486, p-value = 0.03266
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

GRAFIKA 4. TEST JARQUE-BERA

```
Jarque Bera Test

data:  ehat2
X-squared = 7.5549, df = 2, p-value = 0.02288
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

W tym przypadku oba testy nie mają podstawy do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej na poziomie istotności równym 0,01. To oznacza, że jest poprawna specyfikacja modelu oraz normalność rozkładu składnika losowego. Mimo, to zdecydowano się zmodyfikować ten model kolejny raz tak, by był zgodny z modelem w literaturze. Co więcej,

dla modelu 2 zmienna opisująca dochód narodowy kraju pochodzenia nie jest istotna na ustalonym poziomie α równym 0,01.

3.3 Model 3

Ze względu na literaturę, według której należy użyć logarytmu z przepływów migracji oraz z logarytmu dystansu między krajami (Crymble, 2019). Zatem będzie on zgodny z poniższym wzorem:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5,$$

gdzie oznaczenia kolejno znaczą: y – `l_flow`, x_1 – `l_pop_i`, x_2 – `l_pop_j`, x_3 – `l_dist`, x_4 – `l_gni_i`, x_5 – `l_gni_j`. Poniżej przedstawiono wyniki oszacowań modelu 3.

TABELA 3. OSZACOWANIA MODELU 3

```
lm(formula = l_flow ~ l_pop_i + l_pop_j + l_dist + l_gni_i +
    l_gni_j, data = DATA)

Residuals:
    Min       1Q   Median       3Q      Max
-8.2416 -1.7508 -0.1053  1.5810  6.5829

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.27803    1.85918  -6.066 2.63e-09 ***
l_pop_i      0.54065     0.04719  11.457 < 2e-16 ***
l_pop_j      0.47828     0.05330   8.973 < 2e-16 ***
l_dist     -1.39133     0.12203 -11.402 < 2e-16 ***
l_gni_i      0.22191     0.07432   2.986 0.00297 **
l_gni_j      0.96992     0.08271  11.727 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.342 on 488 degrees of freedom
Multiple R-squared:  0.4002,    Adjusted R-squared:  0.394
F-statistic: 65.12 on 5 and 488 DF,  p-value: < 2.2e-16
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Jak widać w tabeli 3 model 3 lepiej opisuje zależność między przepływami migracji a zmiennymi objaśniającymi, gdyż R^2 modelu 3 jest wyższy niż R^2 modelu 2. Co, więcej wszystkie zmienne objaśniające są istotne na poziomie 0,01. Następnie przeprowadzono test RESET i test Jarque-Bera. Poniższe grafiki przedstawiają wyniki testów.

GRAFIKA 5. TEST RESET

RESET test

```
data: model3  
RESET = 3.9468, df1 = 2, df2 = 486, p-value = 0.01994
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

GRAFIKA 6. TEST JARQUE-BERA

Jarque Bera Test

```
data: ehat3  
X-squared = 1.1113, df = 2, p-value = 0.5737
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Oba testy nie mają podstawy do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej przy poziomie istotności wynoszącym 0,01. Zatem, model 3 ma poprawną specyfikację i jego składnik losowy cechuje się normalnością rozkładu. Podjęto decyzję, by przeprowadzić dokładniejszą analizę ekonometryczną modelu 3.

4. Analiza ekonometryczna

4.1 Oszacowania parametrów modelu 3 MNK i ich interpretacja

W modelu 3 współczynnik R^2 wynosi 0,40. To można interpretować, że model jest wyjaśniony w około 40% przez zmienne objaśniające. Oszacowanie współczynnika przy zmiennej l_pop_i , wynosi 0,54. To oznacza, że wzrost populacji kraju pochodzenia o 1% równoznaczny jest wzroście *country_flow*, przepływu migracji między tymi krajami o 0,54%. Z kolei oszacowanie współczynnika zmiennej l_pop_j wynosi 0,47. Analogicznie, wzrost populacji kraju docelowego o 1% znaczy, że jest wzrost przepływu migracji o 0,47%. Estymacja parametru zmiennej l_dist wynosi -1,39. Zatem wzrost dystansu o 1% oznacza spadek przepływu migracji o 1,39%. Oszacowania współczynników zmiennych l_gni_i i l_gni_j wynoszą kolejno: 0,22 i 0,96. To oznacza, że wzrost dochodu narodowego per capita w kraju pochodzenia o 1% równoznaczny jest wzroście przepływu migracji o 0,22%. Podobnie to wygląda dla dochodu narodowego per capita w kraju docelowym, gdzie wzrost o 1% oznacza wzrost przepływów migracji o 0,97%.

4.2 Test RESET

Ponownie przeprowadzono test RESET w cel sprawdzenia poprawności specyfikacji modelu. Wyniki pokazano na grafice 7. Dla przyjętego poziomu istotności równym 0,01, nie ma podstaw do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej, gdyż p-value jest większe niż poziom istotności. Zatem można stwierdzić, że jest poprawna specyfikacja modelu.

GRAFIKA 7. TEST RESET

```
RESET test  
  
data: model3  
RESET = 3.9468, df1 = 2, df2 = 486, p-value = 0.01994
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

4.3 Współliniowość zmiennych objaśniających

Poniższa grafika przedstawia wyniki badania czy istnieje współliniowość między zmiennymi objaśniającymi. Zgodni z nimi, istnieje, a największa korelacja jest dla logarytmu populacji państwa docelowego a logarytmu dochodu narodowego per capita.

GRAFIKA 8. WSPÓLLINIOWOŚĆ

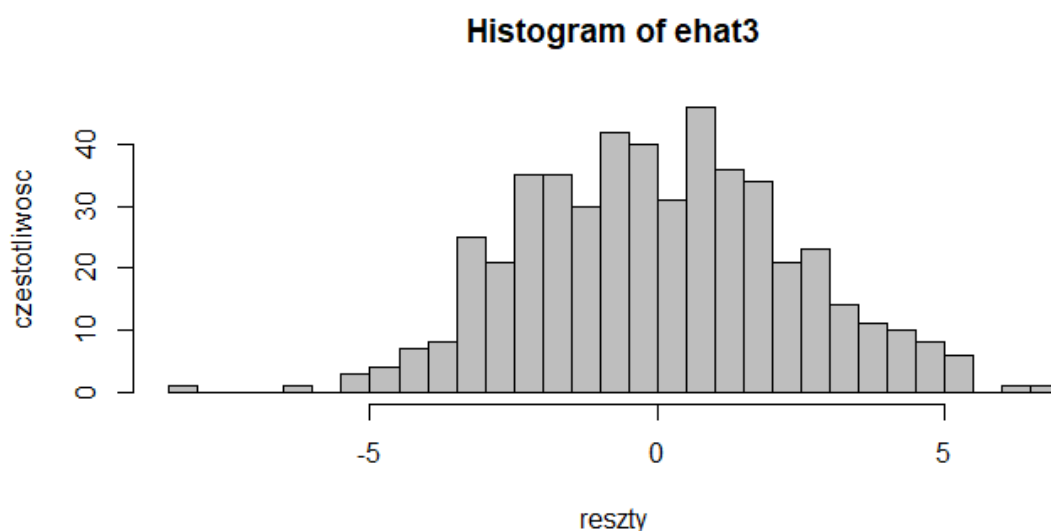
```
> cor(DATA[,
+       c("l_pop_i", "l_pop_j", "l_dist", "l_gni_i", "l_gni_j")
+ ])
      l_pop_i    l_pop_j    l_dist    l_gni_i    l_gni_j
l_pop_i 1.00000000 -0.08052883 0.04806383 -0.105853739 -0.127937526
l_pop_j -0.08052883  1.00000000 0.07353039  0.003955445 -0.239959564
l_dist  0.04806383  0.07353039 1.00000000  0.097564482  0.166011705
l_gni_i -0.10585374  0.003955445 0.09756448  1.000000000  0.002825342
l_gni_j -0.12793753 -0.239959564 0.16601171  0.002825342  1.000000000
> vif(model3)
      l_pop_i    l_pop_j    l_dist    l_gni_i    l_gni_j
1.052169 1.093812 1.063259 1.023614 1.133121
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

4.4 Normalność rozkładu składnika losowego

W celu zbadania własności składnika losowego, zdecydowano się wykonać histogram składnika losowego. Jest on przedstawiony na wykresie 9.

WYKRES 9. HISTOGRAM SKŁADNIKA LOSOWEGO



ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Przedstawiony rozkład przypomina rozkład normalny. Jednak to nie wystarczy do oceny normalności składnika losowego, zatem ponownie przeprowadzono test Jarque-Bera. Przy przyjętym poziomie istotności (0,01), nie ma podstaw do odrzucenia hipotezy zerowej. Hipoteza zerowa dotyczyła normalności składnika losowego.

GRAFIKA 9. TEST JARQUE-BERA

```
Jarque Bera Test  
  
data: ehat3  
X-squared = 1.1113, df = 2, p-value = 0.5737
```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

4.5 Heteroskedastyczność

Następnie zdecydowano się zbadać sferyczność składnika losowego. Wykorzystano do tego dwa testy: test White-a i test Breusch-Pagana. Pierwszy szacuje model regresji względem kwadratu reszt a zmiennymi objaśniającymi i ich interakcjami. Zakłada on, że jeśli model jest istotny to składnik losowy jest heteroskedastyczny. Poniżej przedstawiono wyniki tego testu w grafice 10. Jest on istotny, zatem składnik losowy jest heteroskedastyczny. Warto to jednka potwierdzić przy użyciu drugie testu. Hipoteza zerowa testu Breush-Pagana mówi, że jest obecna homoskedastyczność. Jego wynik widoczny w grafice 11 odrzuca hipotezę zerową przy poziomie istotności równym 0,01, co potwierdza, że składnik losowy charakteryzuje się heteroskedastycznością.

GRAFIKA 10. TEST WHITE-A

```

Residuals:
    Min       1Q   Median       3Q      Max
-11.306  -4.149  -1.619   2.061  52.138

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   43.09875    77.94904   0.553 0.580586
l_pop_i       -4.29961     3.26699  -1.316 0.188785
l_pop_j        8.41775     4.28856   1.963 0.050249 .
l_dist       -16.83052     6.65421  -2.529 0.011752 *
l_gni_i       -2.15493     4.13509  -0.521 0.602518
l_gni_j        0.47046     5.87948   0.080 0.936257
I(l_pop_i^2)    0.01692     0.05025   0.337 0.736421
I(l_pop_j^2)   -0.13076     0.08262  -1.583 0.114160
I(l_dist^2)     0.18974     0.26536   0.715 0.474942
I(l_gni_j^2)    0.12701     0.17711   0.717 0.473653
I(l_pop_i * l_pop_j) -0.04867    0.07176  -0.678 0.497982
I(l_pop_i * l_dist)  0.64996     0.19183   3.388 0.000762 ***
I(l_pop_i * l_gni_i) -0.05072     0.13695  -0.370 0.711269
I(l_pop_i * l_gni_j) -0.03446     0.11161  -0.309 0.757661
I(l_pop_j * l_dist) -0.16266     0.21457  -0.758 0.448794
I(l_pop_j * l_gni_i) -0.04326     0.11287  -0.383 0.701709
I(l_pop_j * l_gni_j) -0.07584     0.17794  -0.426 0.670159
I(l_dist * l_gni_i)  0.43154     0.28780   1.499 0.134430
I(l_dist * l_gni_j)  0.04582     0.32706   0.140 0.888640
I(l_gni_j * l_gni_i) -0.01711     0.18722  -0.091 0.927211
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.996 on 474 degrees of freedom
Multiple R-squared:  0.1252,    Adjusted R-squared:  0.09013
F-statistic:  3.57 on 19 and 474 DF,  p-value: 7.298e-07

```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

GRAFIKA 11. TEST BREUSCH-PAGANA

```

studentized Breusch-Pagan test

data:  model3
BP = 44.241, df = 5, p-value = 2.069e-08

```

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

4.6 Metody uwzględniające niesferyczność składnika losowego

Powyżej potwierdzono niesferyczność składnika losowego, więc należy stworzyć modele, które to uwzględniają. Zastosowano metody, takie jak: odporne błędy standardowe i ważoną MNK w celu uzyskania efektywniejszych oraz nieobciążonych estymatorów. Poniżej przedstawiono tabelę 4, która porównuje oszacowania parametrów dla tych metod.

TABELA 4. PORÓWNANIE MODELU 3 MNK Z ODPORNYMI BŁĘDAMI STANDARDOWYMI I WAŻONĄ MNK MODELU

	Dependent variable:		
	l_flow OLS	coefficient test	l_flow OLS
	MNK (1)	odporne błędy standardowe (2)	ważone MNK (3)
l_pop_i	0.541*** (0.047)	0.541*** (0.048)	0.491*** (0.045)
l_pop_j	0.478*** (0.053)	0.478*** (0.053)	0.479*** (0.052)
l_dist	-1.391*** (0.122)	-1.391*** (0.145)	-1.383*** (0.124)
l_gni_i	0.222*** (0.074)	0.222*** (0.073)	0.227*** (0.071)
l_gni_j	0.970*** (0.083)	0.970*** (0.083)	0.910*** (0.077)
Constant	-11.278*** (1.859)	-11.278*** (1.985)	-10.048*** (1.788)
Observations	494		494
R2	0.400		0.390
Adjusted R2	0.394		0.383
Residual Std. Error (df = 488)	2.342		1.961
F Statistic (df = 5; 488)	65.118***		62.303***
Note: *p<0.1; **p<0.05; ***p<0.01			

ŹRÓDŁO: OPRACOWANIE WŁASNE NA PODSTAWIE ZEBRANYCH DANYCH

Porównując te metody, można zauważyć niewielkie różnice między modelami. Oszacowanie parametrów się niewiele różni między modelem 3 MNK a modelem 3 ważne MNK. Dla wszystkich metod zmienne są istotne statystycznie na poziomie istotności równym 0,01. W tym przypadku użycie MNK może być wystarczające, gdyż nie ma według literatury powody by uwzględnić dodatkowe czynniki sugerujące konieczność uwzględnienia wag.

5. Podsumowanie

Celem niniejszego raportu było przeanalizowanie uzyskanych zależności między przepływami migracji między krajem pochodzenia a krajem docelowym, a populacją tych obu państw i ich GNI *per capita* (Gross National Income *per capita*, po polsku: dochód narodowy *per capita*), i fizycznym dystansem między nimi. Zgodnie z wynikami, można zauważyć, że zostaje potwierdzona teoria. Zgodnie z nią dystans powinien wpływać negatywnie na przepływ migracji i ta zależność została potwierdzona w modelu. Co więcej, pozytywny wpływ wielkości populacji na siłę tego zjawiska również został potwierdzony (Chojnicki, 1966), co jest zgodne z teorią oraz ogólną logiką. Gdyż, jeśli więcej jest osób w miejscu pochodzenia wymusza to na osobach szukania pracy poza tym krajem i jeśli wzrasta populacja w kraju docelowym to może oznaczać, że gospodarka tego kraju pozwala na zatrudnienie większej ilości pracowników. Podobnie to wygląda dla dochodu narodowego *per capita* w kraju docelowym. W literaturze z reguły korzysta się z proporcji między płacami w kraju docelowym a kraju pochodzenia, ponieważ migrant kieruje się największym zyskiem (Anderson, 2011). To może być jednym z ograniczeń tego modelu, gdyż nie zawiera on innej zmiennej, która mogłaby mierzyć atrakcyjność migracji, na przykład język urzędowy w danym kraju, kultura lub nastawienie do migrantów. Dołączenie tych zmiennych warto rozważyć jako potencjalne rozszerzenie badania.

6. Bibliografia

Abel, G.J. and Sander, N. (2014) 'Quantifying Global International Migration Flows', *Science*, 343(6178), pp. 1520–1522. Available at: <https://doi.org/10.1126/science.1248676>.

Anderson, J.E. (2011) 'The Gravity Model', *Annual Review of Economics*, 3(Volume 3, 2011), pp. 133–160. Available at: <https://doi.org/10.1146/annurev-economics-111809-125114>.

Bergstrand, J.H. (1985) 'The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence', *The Review of Economics and Statistics*, 67(3), pp. 474–481. Available at: <https://doi.org/10.2307/1925976>.

Chojnicki, Z. (1966) *Zastosowanie modeli grawitacji i potencjału w badaniach przestrzenno-ekonomicznych*. Wyd. 1. Warszawa: Państwowe Wydawn. Naukowe (Komitet Przestrzennego Zagospodarowania Kraju Polskiej Akademii Nauk. Studia, t. 14).

Conte, M., Cotterlaz, P. and Mayer, T. (2022) 'The CEPII Gravity database' *CEPII Working Paper N°2022-05*. Available at: http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8 (Accessed: 8 June 2024).

Crymble, A. (2019) 'Introduction to Gravity Models of Migration & Trade', *Programming Historian*, 8. Available at: <https://doi.org/10.46430/phen0085>.

'World Migration Report 2024' (2020). Available at: <https://publications.iom.int/books/world-migration-report-2024> (Accessed: 9 June 2024).

World Bank Open Data - GNI per Capita (no date) *World Bank Open Data*. Available at: <https://data.worldbank.org> (Accessed: 8 June 2024).

World Bank Open Data - Population (no date) *World Bank Open Data*. Available at: <https://data.worldbank.org> (Accessed: 8 June 2024).