05.06.2023

„Czy poziom wykształcenia rzeczywiście wpływa na zarobki?”

*Karol Górski 114059*

*Oskar Straszyński 101276*

Spis treści:

* **Cel projektu**
* **Koncepcja modelu**
  1. Dyskusja na temat doboru zmiennych objaśniających
  2. Opis podstawowych statystyk i wykresy
  3. Opis transformacji zmiennych
* **Opis specyfikacji modelu**
  1. Dobór specyfikacji modelu
  2. Sprawdzenie spełnienia założeń MNK
  3. Interpretacja kluczowych parametrów modelu
  4. Weryfikacja własności modelu
* **Podsumowanie**
  1. **Cel projektu**

Niniejszy raport ekonometryczny został zainspirowany pracą Karoliny Goraus-Tańską i Zuzanny Osiki artykułu “Stopnie wyższego wykształcenia, a zarobki w Polsce i innych krajach OEC” (2020). Stopień wykształcenia często służy jako miernik produktywności pracownika, wpływając na poziom życia i pozycję społeczną. Zgodnie z wynikami tego badania osoby posiadające dyplom studiów magisterskich zarabiają więcej niż te z tytułem licencjata we wszystkich krajach uwzględnionych w badaniu. Zatem zwiększanie liczby ludności z wyższym wykształceniem będzie zwiększało średnią stawkę godzinową za roboczogodzinę pracy.

Na ulicach miast wśród młodych ludzi co raz częściej możemy usłyszeć, że studia to strata czasu. Nie raz słyszymy, że ktoś „dorobił się” wcale nie spędzając wielu godzin w książkach, na wykładach gromadząc „zbędną wiedzę” … Chcemy na przekór obecnym trendom i przekazom mass-mediów zbadać czy poziom wykształcenia istotnie wpływa na średnią stawkę godzinową za roboczo-godzinę niezależnie od uwarunkowań kraju, którego mieszkańców analiza dotyczy. Stąd zdecydowaliśmy się przeanalizować dane dla 42 krajów aby zbadać rzeczywisty wpływ wynagrodzenia przy innych ścieżkach rozwojowych.

Celem zbudowanego przez nas modelu ekonometrycznego jest zbadanie wpływu poziomu wykształcenia mieszkańców 42 krajów tj. Australia, Austria, Belgium, Canada, Switzerland, Chile, Colombia, Costa Rica, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greece, Hungary, Ireland, Iceland, Israel, Italy, Japan, Korea, Lithuania, Luxembourg, Latvia, Mexico, Netherlands, Norway, New Zealand, Poland, Portugal, Slovak Republic, Slovenia, Sweden, Türkiye, United States, Singapore, Argentine, Egypt na średnią stawkę godzinową netto wyrażoną w dolarach amerykańskich.

Mówiąc wprost postaramy się odpowiedzieć na pytanie „Czy opłaca się uczyć?”

Posłużymy się pytaniem badawczym w postaci „Czy zdobycie wyższego wykształcenia wpływa istotnie statystycznie na poziom zarobków?”

Analizy dokonujemy na podstawie danych z 2015 roku z bazy danych Eurostat i OECD.

* 1. **Koncepcja modelu.**

1. **Dyskusja na temat zmiennych objaśniających.**

Zmienną objaśnianą w proponowanym przez nas modelu jest średnia stawka godzinowa. Poprzez średnią stawkę godzinową rozumiemy sumę udziałów pracy w Dochodzie Narodowym Netto przypadającą na przepracowaną roboczo-godzinę bez zróżnicowania według płci lub też sektora, w którym mieszkańcy pracują (dane pobrane w takim formacie bezpośrednio z bazy OECD).

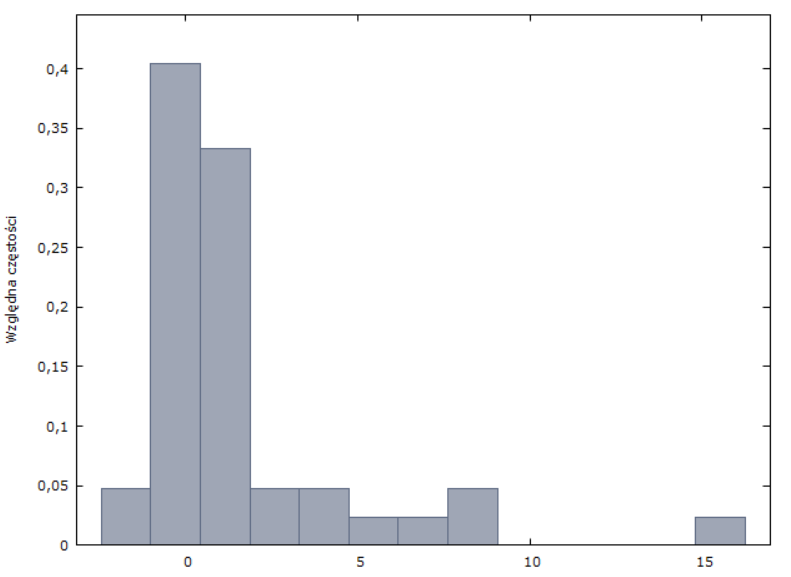
Jako regresory w naszym modelu zaproponowaliśmy zastosowanie poziomu stopy bezrobocia, Produktu Krajowego Brutto, oczekiwanej długości życia i zdobytego wykształcenia. Logicznie każda z zaproponowanych zmiennych objaśniających znacząco wpływa na wysokość zarobków.

* **Produkt Krajowy Brutto** – jest powszechnie uznanym miernikiem poziomu wzrostu gospodarczego, co naturalnie wraz ze wzrostem gospodarczym wzrostowi powinny ulegać średnie stawki płacowe w gospodarce. Zmienna jest wyrażona w dolarach amerykańskich przypadających na 1 mieszkańca.
* **Stopa bezrobocia**- zaproponowaliśmy ją jako jeden z regresorów, gdyż jest bezpośrednio związana z sytuacją na rynku pracy, spodziewamy się tutaj odwrotnej zależność. Mianowicie w przypadku niskiego bezrobocia i braku specjalistów na rynku pracy stawki wynagrodzeń powinny wzrastać. Zmienna wyrażona w procentach.
* **Wskaźnik skolaryzacji**– jest to strategicznie ważna dla naszego modelu zmienna objaśniająca. Wyrażony jest w procentach i reprezentuje odsetek społeczeństwa posiadający wyższe wykształcenie. Intuicja wskazuje na dodatnią zależność pomiędzy poziomem wykształcenia a wysokością średniej płac.
* **Poziom wydatków rządowych na cele edukacyjne** – wydatki rządowe na cele edukacyjne wzmagają rozwój infrastruktury i instytucji edukacyjnych zapewniając powszechny dostęp do edukacji zwiększają produktywność pracy (np. finansowane kursy specjalistyczne). Dane wyrażone jako procent PKB danego kraju.
* **Poziom inflacji** – dodajemy tą zmienną w celu uniknięcia pozornego obrazu wzrostu wynagrodzeń, gdyż ceteris paribus może być on powodowany tylko wzrostem kosztów życia, a nie natomiast innymi strategicznymi determinantami

1. **Omówienie podstawowych statystyk opisowych**

* Obraz zawierający diagram, tekst, zrzut ekranu, Wykres

  Opis wygenerowany automatycznie**Produkt Krajowy Brutto –** Dla analizowanych przez nas krajów Produkt Krajowy Brutto w 2015 roku średnio wyniósł 38 189 USD per capita z wartością minimalną 5580,5 USD per capita i wartością maksymalną 107 900 USD per capita. Gdzie mediana wyniosła 37 575 USD per capita a odchylenie standardowe wyniosło 18 292 USD. Na potrzeby poniższej analizy w modelu posługujemy się wartościami logarytmu naturalnego z wartości PKB w celu precyzyjnego oszacowania parametrów i umożliwienia interpretacji zmian procentowych wartości PKB. Zmienna charakteryzuje się następującym rozkładem częstości (rozkład silnie prawostronny):
* **Stopa bezrobocia -** Dla analizowanych przez nas krajów stopa bezrobocia w 2015 roku średnio wyniosła 1,5411% z wartością minimalną -1,7359% i wartością maksymalną 15,534%. Gdzie mediana wyniosła 0,4814%, a odchylenie standardowe wyniosło 3,23%. Zmienna charakteryzuje się następującym rozkładem częstości (rozkład silnie prawostronny):

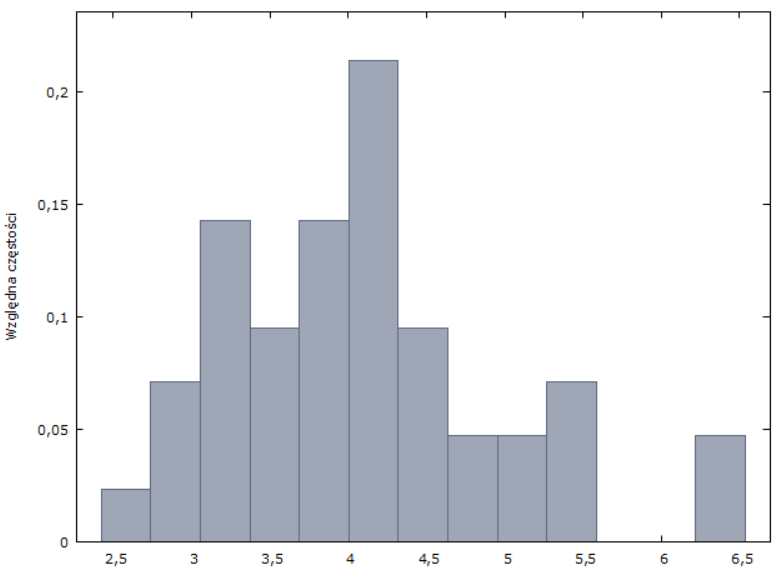


* **Wskaźnik skolaryzacji -** Dla analizowanych przez nas krajów wskaźnik skolaryzacji w 2015 roku średnio wyniósł 33,14% z wartością minimalną 10,58% i wartością maksymalną 55,17%. Gdzie mediana wyniosła 34,58%, a odchylenie standardowe wyniosło 11,24%. Zmienna charakteryzuje się następującym rozkładem częstości:

Obraz zawierający tekst, diagram, zrzut ekranu, Wykres

Opis wygenerowany automatycznie

* **Poziom wydatków rządowych na cele edukacyjne -** Dla analizowanych przez nas krajów procentowy udział wydatków rządowych na cele edukacyjne w 2015 roku średnio wyniósł 4,11% z wartością minimalną 2,58% i wartością maksymalną 6,38%. Gdzie mediana wyniosła 4,05%, a odchylenie standardowe wyniosło 0,87%. Zmienna charakteryzuje się następującym rozkładem częstości:



* **Poziom inflacji (CPI) -** Dla analizowanych przez nas krajów poziom inflacji mierzony za pomocą wskaźnika Consumer Price Index w 2015 roku średnio wyniósł 1,54% z wartością minimalną 1,74% i wartością maksymalną 15,34%. Gdzie mediana wyniosła 0,38%, a odchylenie standardowe wyniosło 3,23%. Zmienna charakteryzuje się następującym rozkładem częstości:

**Obraz zawierający tekst, zrzut ekranu, diagram, Prostokąt

Opis wygenerowany automatycznie**

1. **Opis transformacji zmiennych**

W naszym modelu regresji liniowej postanowiliśmy przetransformować jedną ze zmiennych - Produkt Krajowy Brutto na mieszkańca (GDPoecd) - poprzez zastosowanie transformacji logarytmicznej. Dlatego w modelu zamiast surowego PKB na mieszkańca, mamy log(GDPoecd). Zastosowanie transformacji logarytmicznej pozwala na uwzględnienie nieliniowej zależności między PKB a średnią stawką godzinową. Logarytm ułatwia interpretacje, ponieważ operujemy zmianami procentowymi zmiennej objaśniającej.

Pozostałe zmienne w naszym modelu, tj. stopa inflacji (InfRate), wskaźnik skolaryzacji (School), poziom wydatków rządowych na cele edukacyjne (EducExpen) oraz stopa bezrobocia (Unempl), pozostawiamy w ich pierwotnej formie, ponieważ uznaliśmy, że ich relacja ze średnią stawką godzinową jest najlepiej opisana w liniowej formie. Ta decyzja wynika z chęci zachowania prostoty i przejrzystości naszego modelu, co jest istotne zarówno z punktu widzenia interpretacji wyników, jak i unikania nadmiernego dopasowania. Ponadto, nasza początkowa analiza danych sugeruje, że wybrane zmienne są odpowiednio skonstruowane do badania naszego pytania badawczego i nie wymagają dodatkowych modyfikacji.

1. **Specyfikacja modelu.**

,gdzie :

AvIncome – średnia stawka godzinowa

logGDP – logarytm naturalny wartości PKB

InfRate – stopa inflacji

Unempl – Stopa bezrobocia

School – współczynnik skolaryzacji

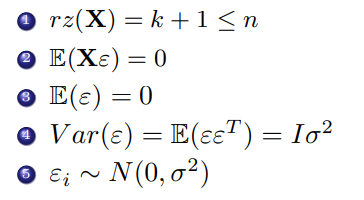
EducExpen – wydatki rządowe na cele edukacyjne

ɛ - składnik losowy

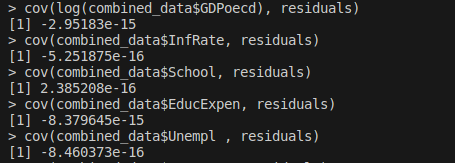
W oparciu o wiedzę ekonomiczną i obserwacje spodziewamy się dodatniej zależności pomiędzy zmienną objaśnianą a zmiennymi: School, logGDP, InfRate, EducExpen i ujemną do zmiennej Unempl.

1. **Analiza ekonometryczna.**

Badając wpływ naszych zmiennych objaśniających na średnią stawkę godzinową chcemy oszacować parametry modelu za pomocą Klasycznej Metody Najmniejszych Kwadratów. Zatem w poniższym punkcie zajmiemy się analizą spełnienia przez zaproponowany przez nas model zestawu warunków w postaci.

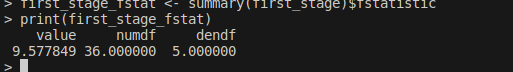


1. Analiza endogeniczności regresorów

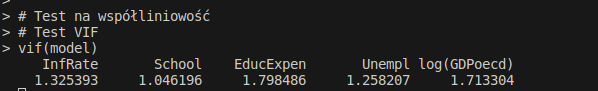


Chcąc posługiwać się Klasyczną Metodą Najmniejszych Kwadratów musimy zbadać czy zastosowane przez nas regresory są egzogeniczne. Niespełnienie tego warunku skutkowałoby obciążonością estymatora MNK i brakiem jego zgodności. Aby dokonać sprawdzenia wykonaliśmy obliczenia kowariancji pomiędzy kolejnymi zmiennymi objaśniającymi a składnikiem losowym z naszego modelu. Powyższe wartości są kowariancjami między składnikiem losowym a zmiennymi objaśniającymi i są bardzo bliskie zeru. Można to interpretować jako brak dowodów na endogeniczność zmiennych objaśniających w modelu, co sugeruje, że nasz model jest dobrze określony i nie ma problemu z endogenicznością. Zatem nie musimy korzystać z alternatywnej metody estymacji jaką jest Metoda Zmiennych Instrumentalnych, gdyż estymator MNK jest zgodny i nieobciążony.

1\* Model oszacowany MZI

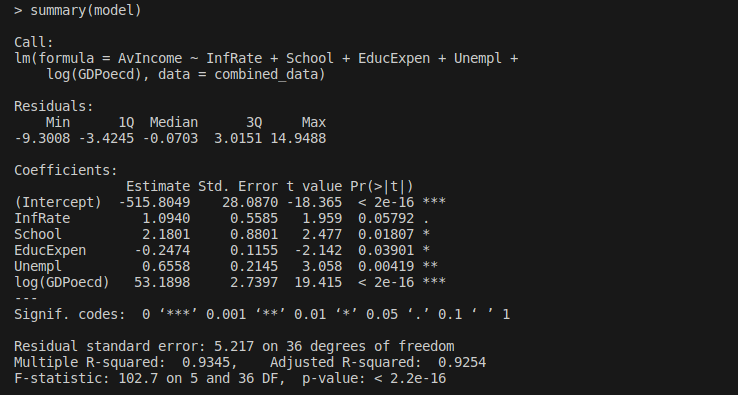


2. Ocena współliniowości zmiennych objaśniających



Wartości Variance Inflation Factor (VIF) powyżej 5 lub 10 są często uznawane za oznakę istotnej współliniowości pomiędzy regresorami. W przypadku naszego modelu, najwyższa wartość VIF to 1.798486 (dla zmiennej EducExpen), co jest znacznie poniżej tych progów. Zatem na podstawie wyników testu VIF, nie wystarczających dowodów na współliniowość w naszym modelu. Zatem możemy stwierdzić, nasz model spełnia będzie najefektywniejszy ze względu na brak współliniowości i niekonieczne jest korzystanie z metody regresji grzbietowej. Estymator KMNK ma korzystniejsze wartości.

3. Ocena łącznej istotności modelu, dopasowania do danych empirycznych i istotności poszczególnych zmiennych objaśniających



Istotność zmiennych objaśniających:

* Zmienna InfRate nie jest istotna na poziomie istotności 0.05 (p-wartość = 0.05792), ale jest bliska do tego. Oznacza to, że jest pewne powiązanie między wskaźnikiem inflacji a średnią stawką godzinową, ale jest ono słabe.
* Zmienna School jest statystycznie istotna na poziomie istotności 0.05 (p-wartość = 0.01807), co sugeruje, że poziom odsetka społeczeństwa posiadającego wyższe wyksztalcenie ma istotny statystycznie wpływ na średnią stawkę godzinową.
* Zmienna EducExpen jest statystycznie istotna na poziomie istotności 0.05 (p-wartość = 0.03901), co wskazuje, że wydatki rządowe na cele edukacyjne mają istotny wpływ na średnią stawkę godzinową.
* Zmienna Unempl jest statystycznie istotna na poziomie istotności 0.05 (p-wartość = 0.00419), co sugeruje, że stopa bezrobocia ma istotny wpływ na średnią stawkę godzinową.
* Zmienna log(GDPoecd) jest wysoce istotna statystycznie (p-wartość < 2e-16), co wskazuje na silny związek między logarytmem PKB na mieszkańca a średnią stawką godzinową.

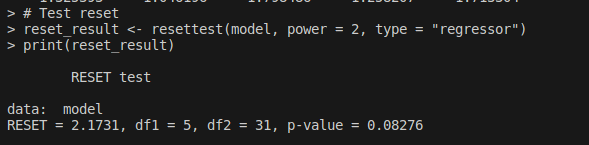
Dopasowanie do danych empirycznych:

* Wskaźnik R-kwadrat wynosi 0.9345, co oznacza, że model wyjaśnia 93,45% wariancji zmiennej zależnej. Jest to bardzo wysoka wartość, co sugeruje, że model dobrze pasuje do danych.
* Dostosowany R-kwadrat (który uwzględnia liczbę regresorów modelu) wynosi 0.9254, co nadal jest wysoką wartością, co wskazuje na dobre dopasowanie modelu.

Łączna istotność modelu (uogólniony test Walda):

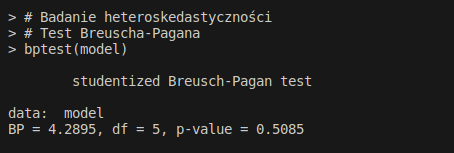
* Statystyka F wynosi 102.7, a wartość p dla testu F jest < 2.2e-16, co jest znacznie poniżej poziomu istotności 0.05. Oznacza to, że model jest istotny, czyli przynajmniej jedna ze zmiennych objaśniających jest istotna.

4. Ocena poprawności specyfikacji modelu

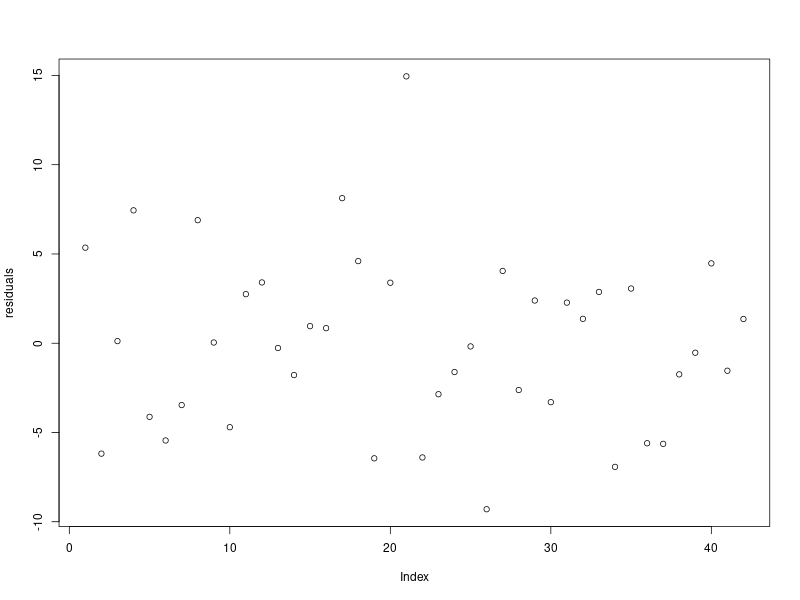


W naszym przypadku wartość statystyki testu RESET wynosi 2.1731. P-wartość dla tego testu wynosi 0.08276, która jest powyżej poziomu istotności 0.05. Oznacza to, że nie mamy wystarczających dowodów statystycznych, aby odrzucić hipotezę zerową, która mówi, że model jest właściwie specyfikowany.

5. Sprawdzenie sferyczności macierzy wariancji-kowariancji składnika losowego – test na występowanie heteroskedastyczności.

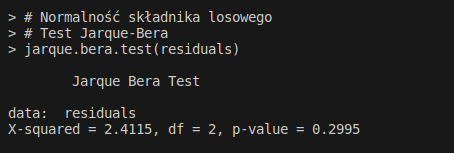


Wartość p wynosi 0,5085, czyli jest większa niż konwencjonalne poziomy istotności (0,1, 0,05 lub 0,01). W ten sposób nie mamy podstaw, aby odrzucić hipotezę zerową o homoskedastyczności składnika losowego. Sugeruje to, że w naszych danych nie ma wystarczających dowodów, aby twierdzić, że wariancje składnika losowego w naszym modelu nie są stałe. Co spełnia kolejny warunek zastosowania KMNK jako najefektywniejszego estymatora.



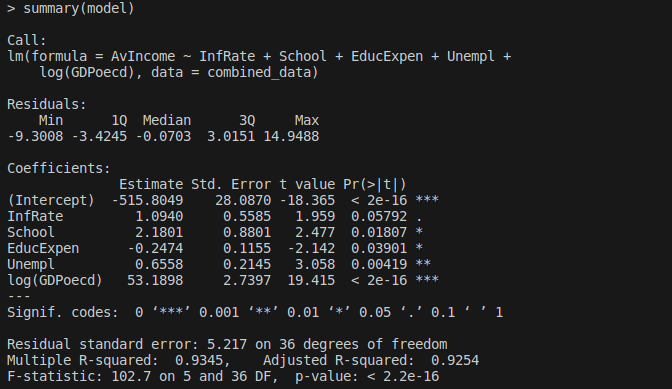
Na powyższym wykresie reszt z modelu nie widać żadnego klarującego się wzoru – ocena organoleptyczna heteroskedastyczności.

6. Sprawdzenie właściwości statystycznych składnika losowego – Test na normalność rozkładu



Dla przeprowadzonego testu Jarque-Bera wartość statystyki testu wynosi 2.4115. P-wartość dla tego testu wynosi 0.2995, co jest powyżej przyjętego poziomu istotności 0.05. To oznacza, że nie mamy wystarczających dowodów statystycznych, aby odrzucić hipotezę zerową, która zakłada, iż składnik losowy ma rozkład normalny. Zatem możemy korzystać z pozostałych testów służących do badania własności składnika losowego, iż są one wiarygodne.

7. Interpretacja oszacowań parametrów modelu



Interpretacja oszacowań parametrów estymatora MNK:

* Constant = -515,80 – w tym przypadku wartość stałej jest pozbawiona logicznej interpretacji
* Oszacowanie parametru dla zmiennej InfRate = 1,09 – wraz ze wzrostem stopy inflacji o 1 p. procentowy średnia stawka godzinowa w danym kraju wzrasta o 1.09 dolara amerykańskiego (Jednakże zmienna nie istotna statystycznie)
* Oszacowanie parametru dla zmiennej School = 2,18 – wraz ze wzrostem odsetka społeczeństwa posiadającego wyższe wykształcenie średnia stawka godzinowa wzrasta przeciętnie o 2,18 dolara amerykańskiego za godzinę.
* Oszacowanie parametru dla zmiennej EducExpen = -0,25 – wzraz ze wzrostem wydatków rządowych na cele edukacyjne w stosunku do PKB o jeden punkt procentowy średnia stawka godzinowa przeciętnie maleje o 0,25 dolara amerykańskiego.
* Oszacowanie parametru dla zmiennej Unempl = 0,65 – wraz ze wzrostem stopy bezrobocia o jeden punkt procentowy średnia stawka godzinowa rośnie o 0,65 dolara amerykańskiego za godzinę pracy.
* Oszacowanie parametru dla zmiennej logGDPoecd = 53,19 – wraz ze wzrostem Produktu Krajowego Brutto o 1% średnia stawka godzinowa wzrasta przeciętnie o 53,19 dolary za godzinę.

**Podsumowanie**

Raport ten podjął się badania wpływu różnych czynników na średnią stawkę godzinową netto (AvIncome) w oparciu o dane OECD i Eurostatu. Model regresji liniowej wykorzystany do analizy zawierał pięć zmiennych objaśniających: stopę inflacji (InfRate), poziom wykształcenia społeczeństwa (School), wydatki rządowe na edukację (EducExpen), stopę bezrobocia (Unempl) oraz logarytm PKB na mieszkańca (log(GDPoecd)). Oszacowania parametrów strukturalnych zostały wykonane za pomocą metody KMNK. Przeprowadzane testy statystyczne wskazały, iż zaproponowane przez nas zmienne wszystkie są egzogeniczne (zbędnym okazało się wykorzystanie Metody Zmiennych Instrumentalnych), ponadto test VIF wskazał, że nie występuje współliniowość regresorów, test RESET wskazał, że nie mamy podstaw, aby odrzucić hipotezę zerową o poprawnej specyfikacji modelu. Dodatkowo test Breush-Pagan'a wskazał, że nie mamy podstaw, aby odrzucić hipotezę zerową o homoskedastyczności składnika losowego oraz test Jarque-Bera wskazał, że składnik losowy ma rozkład normalny. Z powyższego wnioskujemy, iż estymator KMNK jest najefektywniejszy i nieobciążony.

Model wykazał wysokie dopasowanie (R-kwadrat wynosił 0.9345), co sugeruje, że prawie 94% zmienności w średniej stawce godzinowej netto może być wyjaśnione za pomocą pięciu zmiennych objaśniających w modelu. Ponadto zebrane dane i zaproponowany model spełniają założenia twierdzenia Gaussa-Markowa, oznacza to, że estymator Metody Najmniejszych Kwadratów jest BLUE (Best Linear Unbiased Estimator)

Wyniki analizy sugerują, że wszystkie te czynniki mają istotny wpływ na średnią stawkę godzinową netto. Wbrew oczekiwaniom, wydatki rządowe na edukację wykazały ujemny wpływ na średnią stawkę godzinową, co może sugerować, że większe wydatki na edukację niekoniecznie przekładają się na wyższe płace. Tak samo stopa bezrobocia ma zaskakująco pozytywny wpływ na wyższe płace. Natomiast stopa bezrobocia, stopa inflacji, poziom wykształcenia oraz PKB na mieszkańca miały oczekiwany, pozytywny wpływ na średnią stawkę godzinową.

Zmienna log(GDPoecd) miała największy wpływ na średnią stawkę godzinową netto, co sugeruje, że wzrost gospodarczy jest kluczowym czynnikiem wpływającym na płace. Zgodnie z tym, zalecałoby się prowadzenie polityki gospodarczej sprzyjającej wzrostowi PKB na mieszkańca.

Na podstawie zgromadzonych danych i przeprowadzonej analizy można wywnioskować, że poziom wykształcenia społeczeństwa - reprezentowany odsetkiem osób posiadających wyższe wykształcenie, wymiernie dodatnio wpływa na średnią stawkę godzinową w danym kraju.

**Bibliografia**

* Karolina Goraus-Tańska, Zuzanna Osika “Stopnie wyższego wykształcenia a zarobki w Polsce  
  i innych krajach OEC” (2020) https://archiwum.pte.pl/pliki/1/8905/Ekonomista\_2020-3\_7-37.pdf