

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Aleksandra Jędrych
Nr albumu: 419458

**Czy COVID-19 sprawił, że nie chcemy rodzić dzieci?
Analiza determinant współczynnika urodzeń**

Praca wykonana pod kierunkiem
dr Olgi Zajkowskiej
z Katedry Statystyki i Ekonometrii
WNE UW

Warszawa, styczeń 2022

Streszczenie

Pandemia COVID-19 spowodowała, że wiele ludzi musiało zmienić swoje dotychczasowe życie. Zmiany w strukturach demograficznych i ekonomii oddziałują na współczynnik urodzeń. Praca podejmuje problematykę wpływu zmiennych ekonomicznych na podejmowanie decyzji dotyczących prokreacji w powiatach w Polsce.

Słowa kluczowe

współczynnik urodzeń, dzietność, COVID-19, pandemia

Dziedzina pracy (kody wg programu Erasmus)

Ekonomia (14300)

Tytuł pracy w języku angielskim

*Has COVID-19 made us not want to have children?
Analysis of birth rate determinants*

SPIS TREŚCI

WSTĘP.....	4
ROZDZIAŁ I. Przegląd literatury	5
ROZDZIAŁ II. Hipotezy badawcze	8
2.1. Hipotezy o wpływie pandemii COVID-19 na współczynnik urodzeń	8
2.2. Pozostałe hipotezy	9
ROZDZIAŁ III. Analiza danych	10
3.1. Źródła danych	10
3.2. Opis zmiennych	11
3.2.1. Zmienna objaśniana.....	11
3.2.2. Zmienne objaśniające	11
3.3. Statystyczna analiza wybranych zmiennych	13
3.3.1. Zmienna objaśniana – współczynnik urodzeń (<i>birth_rate</i>)	13
3.3.2. Zmienne objaśniające	15
ROZDZIAŁ IV. Formy funkcyjne	20
ROZDZIAŁ V. Diagnostyka modelu	22
5.1. Założenie o prawidłowej formie funkcyjnej	22
5.2. Założenie o homoskedastyczności reszt ε	23
5.3. Założenie o normalności rozkładu reszt ε	25
5.4. Założenie o braku autokorelacji czynników losowych ε	28
5.5. Współliniowość	28
5.6. Obserwacje nietypowe i błędne	29
5.6.1. Dźwignia (<i>leverage</i>)	29
5.6.2. Standaryzowane reszty	30
5.6.3. Odległość Cooka	31
ROZDZIAŁ VI. Weryfikacja hipotez	32
6.1. Interpretacja wyników estymacji	32
6.2. Weryfikacja hipotez badawczych	34

ZAKOŃCZENIE	335
-------------------	-----

WSTĘP

Jednym z problemów, które dotyczą państwa na świecie, w tym Polskę, jest proces intensywnego starzenia się społeczeństwa przy jednoczesnym zmniejszaniu młodszych grup wiekowych. Zjawisko to w szczególności dotyczy kraje rozwinięte i jest następstwem wystąpienia wielu czynników ekonomicznych, społecznych, cywilizacyjnych i kulturowych. Duży wpływ ma między innymi spadek tempa przyrostu naturalnego, na który oddziałuje liczba urodzeń mierzona między innymi przy pomocy współczynnika urodzeń (stosunku liczby żywych urodzeń na 1000 mieszkańców).

Współczynnik urodzeń jest jedną z najbardziej podstawowych, ale również najważniejszych miar demograficznych jak i ekonomicznych. Jego fluktuacje znacząco oddziałują na prowadzenie polityki publicznej przez państwo, w szczególności w obszarach takich jak: edukacja, zdrowie czy system emerytalny.

Celem pracy jest określenie istotności wybranych determinantów współczynnika urodzeń w Polsce. Analiza została stworzona w oparciu o dane dotyczące poszczególnych podregionów Polski – to jest powiatów i miast na prawach powiatu (łącznie 380 obserwacji). W modelu zostały wykorzystane zmienne odnoszące się do kategorii ekonomicznych, społecznych i demograficznych.

Do stworzenia modelu ekonometrycznego wykorzystano dane z 2020 roku pochodzące z Banku Danych Lokalnych (zarządzanego przez Główny Urząd Statystyczny) oraz baz danych gromadzących statystyki dotyczące pandemii COVID-19. W procesie analizy i weryfikowania hipotez badawczych użyto Metody Najmniejszych Kwadratów (MNK).

W pierwszej części pracy został zaprezentowany problem dotyczący niskiej liczby urodzeń oraz dotychczasowe badania związane z tematem modelu. Następnie sformułowano hipotezy badawcze na podstawie dostępnej literatury. W kolejnych częściach uwagę poświęcono zmiennym użytym w modelu, formie funkcyjnej oraz jej diagnostyce. Pracę wieńczą otrzymane wyniki regresji liniowej oraz ich krytyczna analiza.

ROZDZIAŁ I

Przegląd literatury

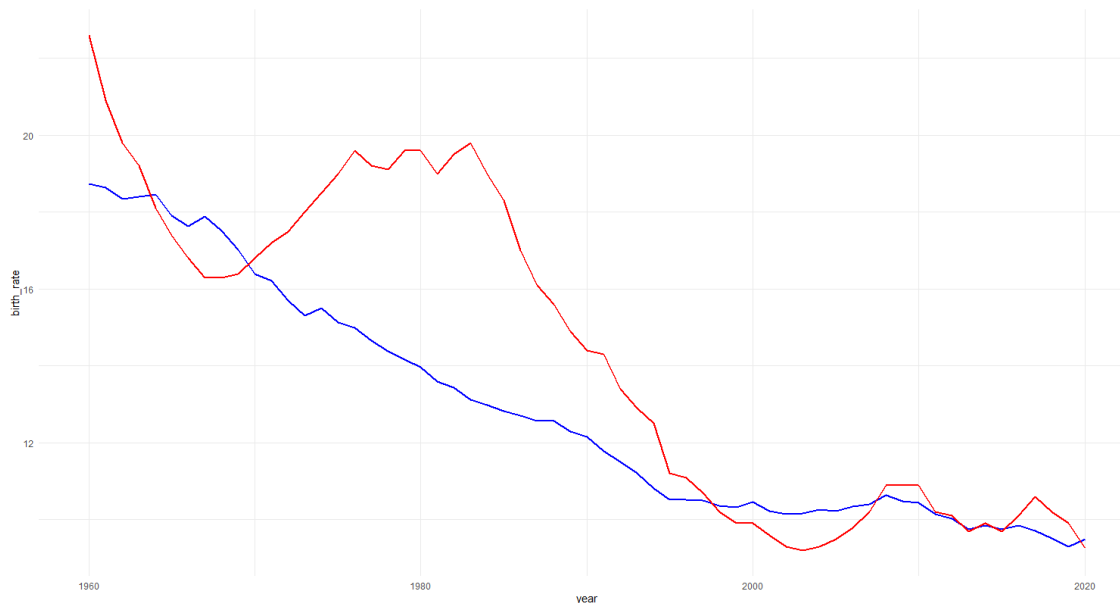
Na przełomie XIX i XX wieku zaczęły powstawać koncepcje dotyczące czynnika ekonomicznego jako stymulatora rozwoju demograficznego. Rozwijały się na płaszczyźnie mikro- i makrospołecznej (pierwsza odnosi się do jednostki, natomiast druga do gospodarstw domowych). W ujęciu makro czynnik ekonomiczny rozpatrywany jest jako teoria dobrobytu, która obejmuje zjawiska ekonomiczne (ilość i jakość dóbr i usług rozporządzana przez społeczeństwo w ramach zaspokajania swoich potrzeb) oraz pozaekonomiczne (związane z fizjologią, psychologią czy etyką). Ówczesni badacze starali się tłumaczyć jak dobrobyt wpływa na liczbę ludności. Perspektywa mikrospołeczna zainteresowała badaczy dopiero w 2. połowie XX wieku w wyniku rozwoju badań nad rodziną i gospodarstwem domowym. Celem badań było zgłębienie wiedzy na temat mechanizmu podejmowania decyzji w sprawie narodzin potomków przez małżeństwa. Na przestrzeni lat powstawały nowe teorie tłumaczące proces, które wiązały się między innymi z kalkulacją użyteczności pochodzącej z dodatkowego dziecka, poszukiwaniem czynników wpływających na częstotliwość współżycia małżonków czy stosowanie metod antykoncepcji. Wynikiem podjęcia próby analizy płodności przez ekonomistów było dostrzeżenie czynników pozabiologicznych, które wpływają na prokreację (Chromińska, 1998).

Problemem w estymowaniu zmiennych dotyczących płodności, w tym współczynnika urodzeń jest to, że w większości koncepcje mają charakter hipotez i trudno zweryfikować je empirycznie. Barięą może być rzetelność odpowiedzi jednostek w badaniach dotyczących tak intymnej sfery jak życie rodzinne. We współczesnych teoriach i analizach Chromińska (1998) dostrzega wadę w postaci pomijania wartości związanych z etyką i moralnością, co wiąże się z niepełnym obrazem oszacowań zmiennych dotyczących wydawania potomstwa na świat. W niniejszej pracy zbadano wpływ warunków ekonomicznych odgrywających dużą rolę w decyzjach dotyczących posiadania dzieci. Pominęto szereg zmiennych, których aspektami zajmuje się medycyna, takimi jak mechanizmy kontroli urodzeń czy terminacje ciąży.

Od początku lat 90. XX wieku w Polsce możemy zaobserwować wyraźny spadek współczynnika urodzeń (Rys. 1.). Według demografów wynika to przede wszystkim ze zmiany poglądów na temat posiadania dziecka przez młodych ludzi – decyzja o potomstwie jest odkładana na później (między innymi z powodu chęci osiągnięcia określonego poziomu

wykształcenia i stabilizacji ekonomicznej). Młodzi decydują się na mniejszą liczbę dzieci lub wolą żyć samotnie. Badacze wskazują, że sytuacja ludnościowa jest trudna, ponieważ niski poziom dzietności negatywnie wpływa na brak zastępowalności pokoleń (konsekwencja zmniejszenia się liczby kobiet w wieku rozrodczym) (Hrynkiewicz, Witkowski, Potrykowska, 2018).

Rys. 1. Wykres zmian współczynnika urodzeń dla Polski i państw Unii Europejskiej w latach 1960-2020.



Źródło: Opracowanie własne na podstawie danych *World Bank* z wykorzystaniem biblioteki *ggplot2*.

Prognozy przedpandemiczne wskazywały, że połączenie niskiego współczynnika urodzeń i wydłużającego się czasu trwania życia może spowodować zmniejszenie podaży pracy a także postępujące starzenie się społeczeństwa. Proces starzenia się ludności jest wyzwaniem, ponieważ wpływa na wiele sfer życia (ekonomiczną, psychologiczną, medyczną czy socjalną). Dodatkowo starość demograficzna oznacza też silną feminizację w grupach starszych wiekowo (ze względu na średnio dłuższe trwanie życia kobiet w stosunku do mężczyzn – spowodowane nawykami żywieniowymi, higienicznymi, warunkami pracy) (Hrynkiewicz, Witkowski, Potrykowska, 2018).

Z perspektywy ekonomicznej starzejące się społeczeństwo może powodować postępujące zmniejszanie się wskaźników wzrostu gospodarczego w regionach. Początkowo, wraz z malejącą liczbą urodzeń gospodarka rośnie, jednak gdy zostanie przekroczony punkt krytyczny a spadek dzietności niezahamowany to równowaga między populacją osób starszych a grupą osób w wieku produkcyjnym nie będzie mogła być utrzymana (Arcuri, 2019).

Konsekwencją procesu będzie potrzeba reorganizacji sposobu finansowania wydatków publicznych, ponieważ coraz więcej osób będzie wymagało zabezpieczenia emerytalno-rentowego (Hryniewicz, Witkowski, Potrykowska, 2018).

Temat wpływu warunków ekonomicznych na podejmowanie decyzji dotyczących prokreacji został podejmowany przez wielu badaczy. Różne zmienne opisujące stan gospodarki w zróżnicowanym stopniu oddziałują na liczbę urodzonych dzieci. Istotną rolę w kontekście urodzeń przypisuje się wynagrodzeniom. Ermisch (1998) w ramach opisu modelu szeregów czasowych tłumaczącego współczynnik urodzeń stwierdza, że wyższa płaca netto w przypadku kobiet zniechęca je do wydawania na świat potomków, natomiast przeciwne wnioski wyciąga dla mężczyzn. Gary Becker w artykule „*An Economic Analysis of Fertility*” również wskazuje dochód jako jedną z istotnych determinant liczby urodzeń. Związek między zarobkami a liczbą urodzeń zauważają też Kearney i Willson (2018). Vandenbroucke (2016) natomiast punktuje negatywną tendencję jeśli chodzi o PKB per capita państw a urodzeniami. Mniejsze wynagrodzenie świadczy o relatywnie tańszym czasie, więc jego poświęcenie na wychowanie dziecka nie jest tak kosztowne jak w przypadku regionów bogatszych (notujących wyższe średnie wynagrodzenie).

Ekonomiści zauważają także zależność między zmianami na rynku mieszkaniowym a współczynnikiem urodzeń. Dettling i Kearney (2014) dowodzą, że wzrost cen mieszkań w krótkim okresie powoduje spadek urodzeń wśród osób nieposiadających nieruchomości oraz wzrost urodzeń pośród właścicieli. Pozytywną korelację między prokreacją a wzrostem cen dla osób posiadających własne domy dowodzą także badania Lovenheima i Mumforda (2013).

Wiele prac wskazuje na istotne powiązanie fluktuacji współczynnika urodzeń i poszczególnych etapów cyklu gospodarczego, w szczególności ze zmianami na rynku pracy. Pandemia COVID-19 wywarła ogromny wpływ na struktury demograficzne państw, jak i na ich gospodarki. Nieprzewidziane w dotychczasowych prognozach wzrost zgonów (Tabela 1.) i zahamowanie wzrostu gospodarczego będą odnotowywane we wskaźnikach ekonomicznych, społecznych czy demograficznych.

Tabela 1. Zestawienie liczby zgonów w Polsce w latach 2017 – 2020

Lata	2017	2018	2019	2020
Liczba zgonów	402 852	414 200	409 709	477 355
Indeks łańcuchowy (rok bazowy = rok poprzedni)	-	103,2%	98,9%	116,5%

Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych*.

Porównując zmniejszenie mobilności obywateli wywołane koronawirusem do sytuacji wprowadzenia stanu wojennego w 1981 roku to należałoby przewidywać wzrost liczby urodzeń spowodowany pozostaniem w domach razem z partnerami. Podobne wnioski zostały formułowane na podstawie historii związanych ze skokiem urodzeń dziewięć miesięcy po zamieciach czy poważnych przerwach w dostawie prądu. Mity te zostały statystycznie obalone między innymi przez Richarda Udry’ego w 1970 roku, który zbadał wpływ awarii zasilania w Nowym Jorku w 1965 roku na urodzenia dzieci. Wzrost współczynnika urodzeń nie został także odnotowany po historycznej zamieci śnieżnej w 1967 roku w Chicago, dlatego można odrzucić hipotezę o zwiększeniu współczynnika współżycia seksualnego prowadzącego do wzrostu liczby urodzeń w przypadku nakazów pozostania w domu (Udry, 1970).

Kryzys zdrowia publicznego wywołał także recesję gospodarczą prowadzącą do strat ekonomicznych, niepewności, braku możliwości przewidywania przyszłości i zmniejszeniu poczucia bezpieczeństwa wobec śmiertelnego wirusa. Badanie przeprowadzone przez Lindeber, VandeVusse, Mueller i Kirstein (2020) ujawnia, że pandemia wpłynęła na zachowania jednostek w obszarze prokreacji – powyżej 40% kobiet zmieniło plany dotyczące potomstwa (ich liczby lub odłożyło ich posiadanie na bardziej sprzyjające ludzkości czasy). COVID-19 wpływając na rynek pracy, śmiertelność, zmniejszenie mobilności obywateli w pośredni sposób wywołuje zmiany w wartości współczynnika urodzeń.

ROZDZIAŁ II

Hipotezy badawcze

2.1. Hipotezy o wpływie pandemii COVID-19 na współczynnik urodzeń

H0: Pandemia COVID-19 negatywnie wpływa na liczbę urodzonych dzieci – ze względu na wzrost bezrobocia spowodowany zamrożeniem gospodarki.

Pandemia wywarła piętno na wielu aspektach życia obywateli. Na skutek lockdownu wiele przedsiębiorstw zbankrutowało, co spowodowało wzrost stopy bezrobocia. Przejściowe zmiany w warunkach ekonomicznych rodzin prowadzą do zmiany liczby urodzeń – wzrost bezrobocia w krótkim okresie wpływa negatywnie na prokreację. Rodziny odkładają narodziny dziecka na później, co w swojej pracy stwierdzają Currie i Schwandt (2014). Lindo (2010) zauważa również, że małżeństwa, w których mężowie stracili w którymś momencie życia pracę mają mniej dzieci.

H1: Pandemia COVID-19 negatywnie wpływa na liczbę urodzonych dzieci – ze względu na śmiertelność.

Oprócz zmian na rynku pracy, pandemia niosła śmiertelne żniwo szczególnie wśród osób starszych. Dziadkowie często współwychowują swoje wnuki i są aktywnie zaangażowani w ich opiekę. Ich śmierć może również przyczynić się do podjęcia decyzji o braku potomstwa. Z tego względu liczba zgonów spowodowanych pandemią wpływa na współczynnik urodzeń.

2.2 Pozostałe hipotezy

H2: Zwiększenie wydatków na dzieci powoduje wzrost współczynnika urodzeń.

Zwiększenie wydatków na dzieci to w przypadku niniejszej pracy wypadkowa zmiennych dotyczących: zwiększenia wydatków budżetów powiatów (i miast na prawach powiatu) na rodziny i pomoc społeczną, której celem jest umożliwienie rodzinom przezwyciężenie

John Ermisch w artykule „*Economic influences on birth rate*” sugeruje, że zachętą do rodzenia dzieci są wprowadzane przez państwo narzędzia polityki fiskalnej takie jak ulgi podatkowe na dzieci lub wypłaty zasiłków. Jednym z takich narzędzi są zasiłki rodzinne, które możemy interpretować jako wydatki budżetów powiatów i miast na prawach powiatu w dziale Rodzina. Wyższe wartości zasiłków (to jest wydatków) wpływają pozytywnie na liczbę rodzonych dzieci (które wpływają bezpośrednio na współczynnik urodzeń).

Bauernschuster, Hener i Rainer (2013) w swoich badaniach wykazali, że rozwój opieki nad dziećmi ma znaczący pozytywny wpływ na prokreację. Prowadzenie polityki prorodzinnej przez państwo, to znaczy ułatwiającej pogodzenie wychowywania dzieci i pracy może być skutecznym sposobem na walkę ze zmniejszającym się współczynnikiem urodzeń.

H3: Wzrost cen mieszkań na rynku ujemnie koreluje ze współczynnikiem urodzeń.

Ermisch (1988) uważa, że konkurencja na rynku mieszkaniowym zniechęca pary do posiadania dzieci. Małżeństwa, które chcą kupić mieszkanie podczas gdy popyt na nie przewyższa podaż będą negatywnie patrzeć na wizję powiększania rodziny w dobie rosnących cen zakupu domu oraz wydłużających się kolejek po mieszkania socjalne. Podobny negatywny efekt cenowy wśród najemców zauważają Dettling i Kearney (2014). Ermisch (1988) jako wskaźnika presji na rynku mieszkaniowym używa średniej ceny nowego domu, w niniejszej pracy zaproponowano mierzyć ten efekt za pomocą mediany ceny za 1 m² lokalu mieszkaniowego.

Z drugiej strony, osoby które posiadają już własne domy w wyniku rosnących cen mieszkań stają się bardziej zamożne, co może być bodźcem do powiększania rodziny. Do innych wniosków dochodzą Black, Kolesnikova, Sanders i Taylor (2013), którzy stwierdzają, że małżeństwa żyjące w bogatych lokalizacjach (czyli w przypadku tej pracy o wyższej medianie cen za m² lokalu mieszkaniowego) statystycznie mają mniejsze rodziny (współczynnik urodzeń jest mniejszy).

ROZDZIAŁ III

Analiza danych

3.1. Źródło danych

Dane przekrojowe wykorzystane w modelu ekonometrycznym dotyczą 380 powiatów i miast na prawach powiatu w Polsce i wyrażają stan na 2020 rok. Do przeprowadzenia estymacji regresji liniowej za pomocą Metody Najmniejszych Kwadratów (MNK) i stworzenia wykresów i tabel znajdujących się w pracy wykorzystano język programowania R. Praca opiera się na danych pochodzących z trzech źródeł: Banku Danych Lokalnych, projektu Michała Rogalskiego – COVID-19 w Polsce oraz archiwalnych danych dla powiatów związanych z koronawirusem COVID-19 udostępnionych przez Ministerstwo Zdrowia.

Wskaźniki i dane dotyczące sytuacji gospodarczej, demograficznej, społecznej powiatów zostały pobrane z Banku Danych Lokalnych, za którego prowadzenie i rozwój odpowiedzialny jest Główny Urząd Statystyczny w Polsce.

Baza danych COVID-19 w Polsce stworzona przez Michała Rogalskiego na podstawie raportów Ministerstwa Zdrowia, Wojewódzkich Stacji Sanitarno-Epidemiologicznych (WSSE), Powiatowych Stacji Sanitarno-Epidemiologicznych (PSSE) i Urzędów Wojewódzkich została wykorzystana do pobrania danych dla powiatów (i miast na prawach powiatu) dotyczących sumy przypadków zakażeń koronawirusem i zgonów nim

spowodowanych. Okres sumowania zgonów jest różny dla różnych województw i zależy od daty zaprzestania publikowania danych dotyczących pandemii przez PSSE i WSSE.

Dane archiwalne dotyczące zarażeń koronawirusem dla powiatów opublikowane przez Ministerstwo Zdrowia (MZ) posłużyły do oszacowania łącznej liczby osób przebywających na kwarantannie w 2020 roku dla poszczególnych powiatów. W archiwum znajdują się statystyki dotyczące COVID-19 w przedziale czasowym od 24.11.2020 do 31.12.2020 roku.

Sposób szacowania łącznej liczby osób przebywających na kwarantannie:

1. Na podstawie danych z okresu 23.11.2020 – 31.12.2020 wyliczono średnią dzienną liczbę osób objętych kwarantanną.
2. Zgodnie z rozporządzeniem Ministra Zdrowia, stan epidemii koronawirusa w Polsce obowiązuje od 20.03.2020 roku. Średnią dzienną liczbę osób objętych kwarantanną wyliczoną w punkcie 1. pomnożono przez liczbę dni objętych stanem epidemii (286 dni) i zaokrąglono do liczb całkowitych.
3. Jeżeli dany powiat nie został uwzględniony w zestawieniu Ministerstwa Zdrowia, to dane dotyczące kwarantann na jego obszarze zostały wyliczone na podstawie średniej dziennej liczby osób objętych kwarantanną dla całego kraju.

3.2. Opis zmiennych

3.2.1. Zmienna objaśniana

Zmienną zależną w szacowanym modelu jest współczynnik urodzeń (*birth_rate*), który określa liczbę żywych urodzeń na 1000 ludności.

$$birth.rate = \frac{liczba\ urodze\acute{n}\ \acute{z}ywych}{1000\ ludno\acute{s}ci}$$

3.2.2. Zmienne objaśniające

Wszystkie zmienne niezależne użyte w modelu są zmiennymi ciągłymi.

- *covid19_cases* - suma przypadków zachorowań na chorobę COVID-19.
- *covid19_deaths* – suma zgonów wywołanych chorobą COVID-19.
- *covid19_quarantine* – suma osób przebywających na kwarantannie z powodu kontaktu z osobą chorą na COVID-19.
- *marriage_rate* – współczynnik małżeństw określający liczbę zawartych małżeństw na 1000 ludności.

$$\text{marriage.rate} = \frac{\text{liczba zawartych małżeństw}}{1000 \text{ ludności}}$$

- *divorce_rate* – współczynnik rozwodów, który jest stosunkiem liczby orzeczonych rozwodów na 1000 ludności.

$$\text{divorce.rate} = \frac{\text{liczba rozwodów}}{1000 \text{ ludności}}$$

- *budget_reve_pc* – dochody budżetów powiatów (lub miast na prawach powiatu) w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.
- *unemployment_rate* – stopa bezrobocia rejestrowanego wyrażona w procentach.
- *women_reproductive* – liczba kobiet w wieku reprodukcyjnym, czyli między 15. a 49. rokiem życia w stosunku do liczby kobiet w danym powiecie (lub mieście na prawach powiatu).
- *femininity_ratio* – współczynnik feminizacji określający liczbę kobiet przypadającą na 100 mężczyzn.

$$\text{femininity.ratio} = \frac{\text{liczba kobiet}}{100 \text{ mężczyzn}}$$

- *avg_salary* – średnie wynagrodzenie brutto. Zmienna wyrażona w złotych polskich.
- *women_working* – odsetek kobiet pracujących jako stosunek liczby pracujących kobiet do liczby kobiet w danym powiecie.

$$\text{women.working} = \frac{\text{liczba pracujących kobiet}}{\text{liczba kobiet}}$$

- *men_working* – odsetek mężczyzn pracujących jako stosunek liczby pracujących mężczyzn do liczby mężczyzn w powiecie.

$$\text{men.working} = \frac{\text{liczba pracujących mężczyzn}}{\text{liczba kobiet}}$$

- *median_house_price* – mediana cen za 1 m² lokali mieszkaniowych sprzedanych w ramach transakcji rynkowych. Zmienna wyrażona w złotych polskich.
- *house_ratio* – mieszkania oddane do użytkowania na 1000 ludności.
- *houses_area_pc* – stosunek powierzchni użytkowej mieszkań w danym powiecie do jego populacji. Zmienna wyrażona w metrach kwadratowych.

$$\text{houses.area.pc} = \frac{\text{łączna powierzchnia użytkowa mieszkań}}{\text{populacja}}$$

- *avg_people_per_house* – średnia liczba osób przypadająca na jedno mieszkanie.

$$\text{avg.people.per.house} = \frac{\text{populacja}}{\text{liczba mieszkań}}$$

- *education_expenditure* - wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 801 – Oświata i wychowanie w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.

$$education_expenditure = \frac{\text{wydatki na oświatę ogółem}}{\text{liczba ludności}}$$

- *health_expenditure* - wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 851 – Ochrona zdrowia w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.

$$health_expenditure = \frac{\text{wydatki na ochronę zdrowia ogółem}}{\text{liczba ludności}}$$

- *social_expenditure* – wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 852 – Pomoc społeczna w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.

$$social_expenditure = \frac{\text{wydatki na pomoc społeczną ogółem}}{\text{liczba ludności}}$$

- *family_expenditure* – wydatki budżetów powiatów (i miast na prawach powiatu) w dziale 855 – Rodzina w przeliczeniu na jednego mieszkańca. Zmienna wyrażona w złotych polskich.

$$family_expenditure = \frac{\text{wydatki na rodzinę ogółem}}{\text{liczba ludności}}$$

- *children* – liczba osób w przedziale wiekowym 0 – 2 w danym powiecie.
- *nursery_places* – miejsca w żłobkach, oddziałach żłobków i klubach dziecięcych.
- *doctors* – personel pracujący w sektorze ochrony zdrowia na 10 tysięcy ludności.
- *urbanisation_rate* – współczynnik urbanizacji określający udział mieszkańców miast w liczbie ludności powiatu (lub miasta na prawach powiatu).
- *bus_stops* – liczba przystanków autobusowych (z trolejbusowymi) i tramwajowych.

3.3. Statystyczna analiza wybranych zmiennych

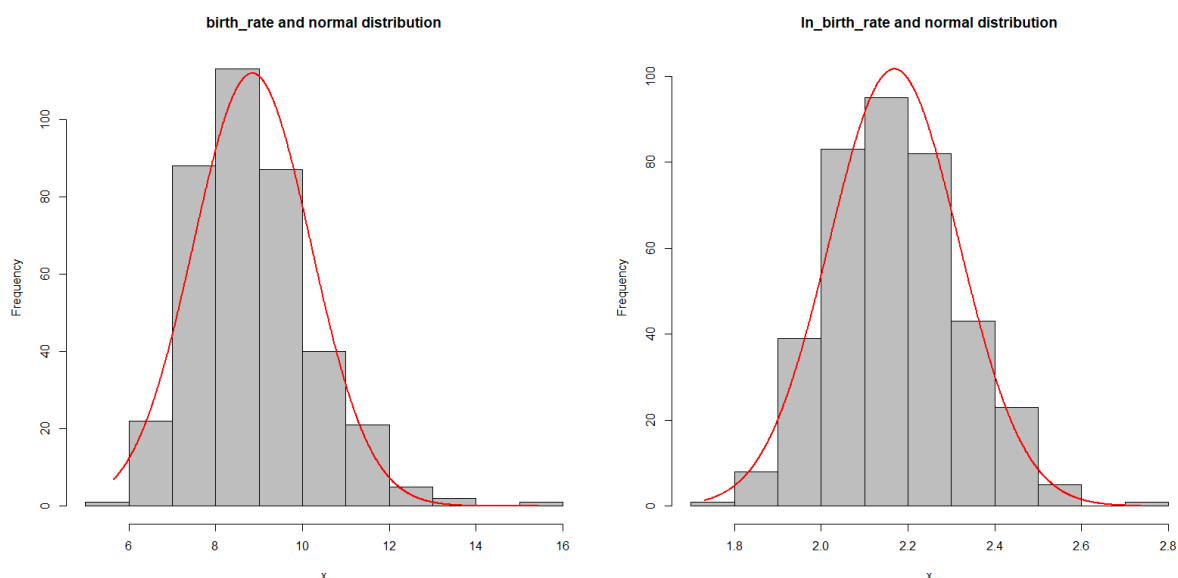
3.3.1. Zmienna objaśniana – współczynnik urodzeń (*birth_rate*)

Rysunek 2. przedstawia histogramy zmiennej objaśnianej modelu ekonometrycznego w wersji podstawowej *birth_rate* oraz zlogarytmowanej *ln_birth_rate*. Zmienna przyjmuje wartości od 5,45 do 15,43 (co możemy odczytać z tabeli z podstawowymi statystykami zmiennych umieszczonej w Załączniku 1.), natomiast najczęściej odnotowywane wartości

znajdują się w przedziale od 8 do 9. Oba histogramy są zbliżone do rozkładu normalnego, dlatego zmienna zależna w wersji podstawowej i zlogarytmowanej została przetestowana pod względem dopasowania do rozkładu normalnego za pomocą testu Jarque-Bera. Wyniki przeprowadzonych testów wskazały, że nie ma podstaw do odrzucenia hipotezy zerowej stanowiącej o rozkładzie normalnym dla zmiennej w postaci zlogarytmowanej – \ln_birth_rate (p-value = 0.06457). W przypadku zmiennej $birth_rate$ hipoteza została odrzucona. Zatem w modelu wykorzystano zmienną \ln_birth_rate .

Warto zaznaczyć, że Dettling i Kearney (2014) w swoim równaniu regresji używają logarytmu ze współczynnika dzietności jako zmiennej objaśnianej, dlatego warto wykorzystać tę myśl i użyć podobnego rozumowania w niniejszej pracy w kontekście modelowania logarytmu współczynnika urodzeń.

Rys. 2. Histogramy współczynnika urodzeń (zmienna bez logarytmu – $birth_rate$ i zmienna zlogarytmowana – \ln_birth_rate)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

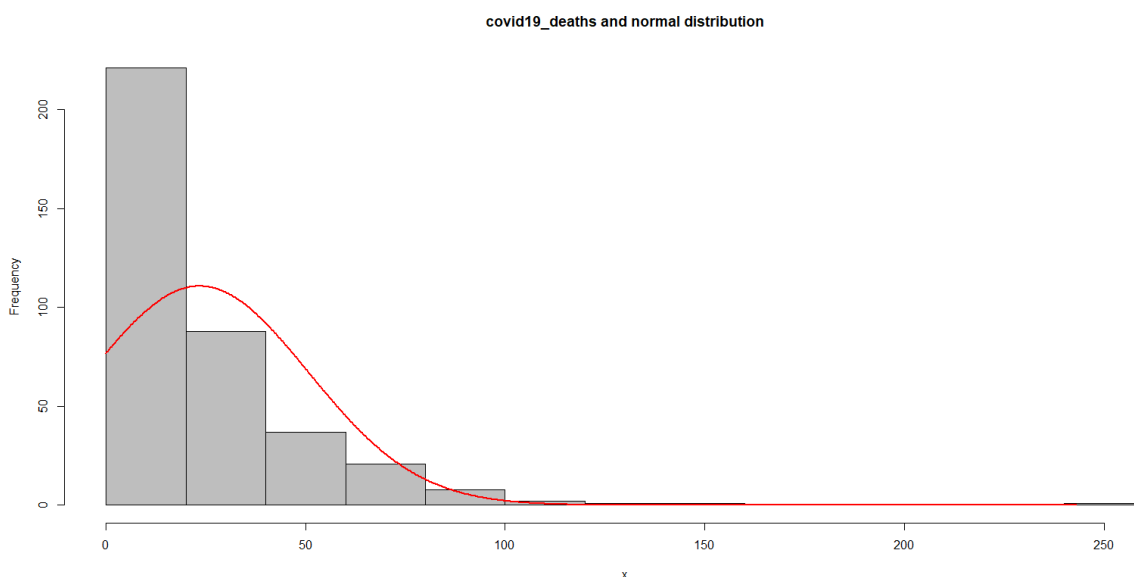
3.3.2. Zmienne objaśniające

Zmienna opisująca pandemię COVID-19 (*covid19_deaths*)

Jako jedną ze zmiennych objaśniających wpływ pandemii COVID-19 na współczynnik urodzeń zaproponowano sumę przypadków zachorowań na chorobę wywołaną wirusem SARS-CoV-2 w danym powiecie (lub mieście na prawach powiatu).

Na podstawie histogramu (Rys. 3.) można zaobserwować, że mediana zmiennej znajdują się w przedziale między 0 a 18.75, a dokładniej wynosi 15. Średnia liczba zgonów w badanej próbie to 23.38. W przypadku zmiennej *covid19_deaths* obserwacje przyjmują wartości 0, a różnica między minimalną a maksymalną wartością wynosi 243. Może być to spowodowane zjawiskiem związanym z niedoszacowaniem wartości dla niektórych powiatów (lub miast na prawach powiatu).

Rys. 3. Histogram zmiennej: *covid19_deaths*



Źródło: Opracowanie własne na podstawie danych *Ministerstwa Zdrowia* z wykorzystaniem biblioteki *rcompanion*.

W celu zbadania normalności rozkładu zmiennej związanej ze zgonami spowodowanymi wystąpieniem choroby COVID-19 przeprowadzono test Jarque-Bera. Przy założonym poziomie istotności $\alpha = 0.05$, hipoteza zerowa o rozkładzie normalnym zmiennej została odrzucona.

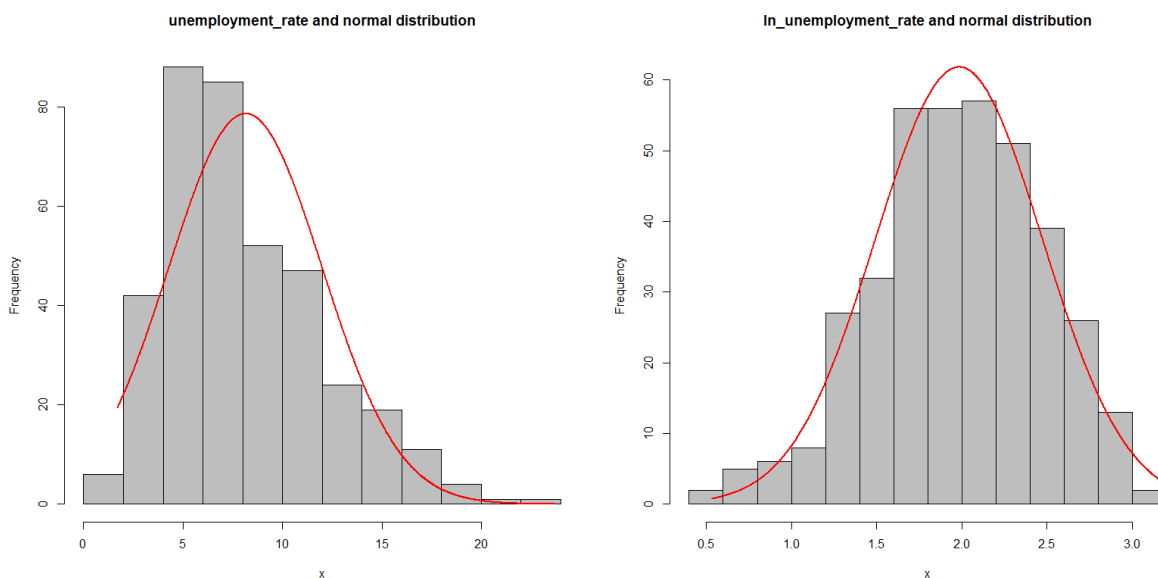
Stopa bezrobocia rejestrowanego (*unemployment_rate*)

John Ermisch (1988) w swoim modelu zastosował logarytm ze zmiennej dotyczącej stopy bezrobocia wśród mężczyzn. W szacowanym modelu wykorzystano pomysł i zaproponowano logarytmowanie zmiennej mierzącej stopę bezrobocia rejestrowanego w każdym z powiatów (i miast na prawach powiatu).

Na podstawie histogramów (Rys. 4.) możemy zauważyć, iż zaproponowana zlogarytmowana zmienna w większym stopniu przypomina zmienną o rozkładzie normalnym. Sprawdzając przypuszczenia za pomocą testu Jarque-Bera okazuje się, że w przypadku obu wersji odrzucamy hipotezę o normalności rozkładu, jednak dla *ln_unemployment_rate* p-value jest znacznie wyższe i prawie osiąga zakładany poziom istotności $\alpha = 0.05$, ponieważ wynosi 4.807%.

Jeżeli przyjrzymy się histogramowi dla stopy bezrobocia w powiatach, to dla większości przyjmuje wartość znajdującą się w przedziale od 4 do 8 procent. Zmienna ma rozkład asymetryczny prawostronny, a jej najwyższa wartość to 23.7%.

Rys. 4. Histogram dla zmiennej opisującej stopę bezrobocia rejestrowanego w wersji podstawowej (*unemployment_rate*) oraz zlogarytmowanej (*ln_unemployment_rate*)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

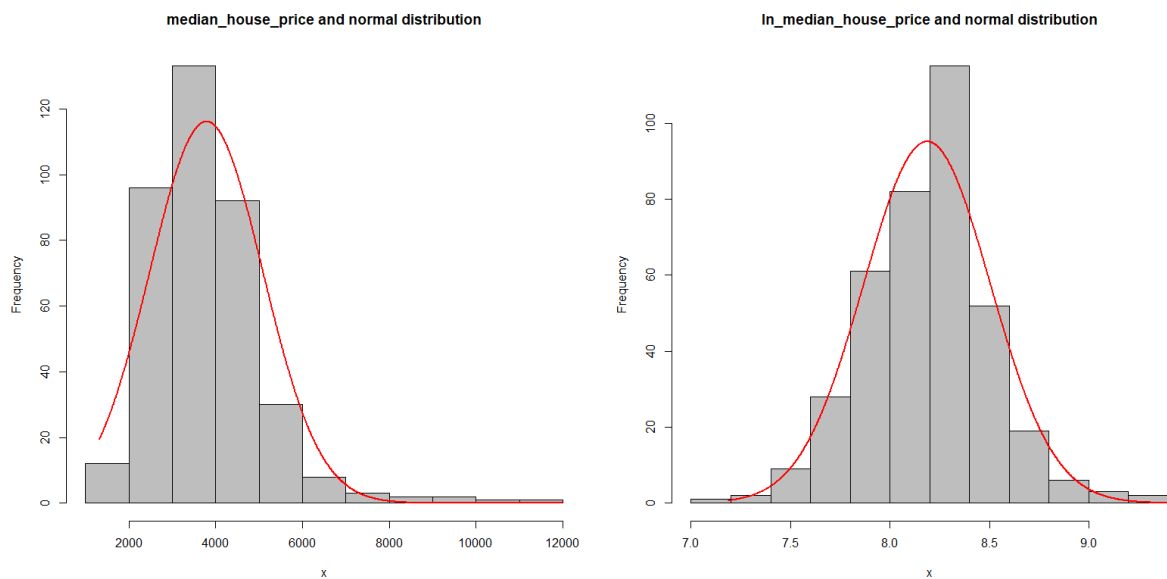
Zmienna opisująca rynek mieszkaniowy (*median_house_price*)

Jako zmienną opisującą rynek mieszkaniowy zaproponowano medianę cen za 1 m² lokali mieszkaniowych sprzedanych w ramach transakcji rynkowych.

Na podstawie histogramu (Rys. 5.) dla zmiennej *median_house_price* możemy wnioskować, że w większości powiatów najczęściej występująca cena za metr kwadratowy lokalu mieszkaniowego mieści się w przedziale od 3000 zł do 4000 zł. Średnia wartość zmiennej wynosi 3786,32 zł. Rozstrzał cenowy jeśli chodzi o mieszkania jest duży, ponieważ różnica między maksymalną a minimalną wartością *median_house_price* wynosi 10 663 złote. Najwyższe ceny notuje się w miastach, na przykład według strony muratorplus.pl przeciętna cena mieszkania w I półroczu 2020 roku wynosiła 10 800 zł za m².

John Ermisch (1998) w swoim modelu szacującym wartość współczynnika urodzeń jako jedną ze zmiennych proponuje logarytm z realnej ceny domu. Chcąc wykorzystać myśl badacza w niniejszej pracy, zaproponowano użycie logarytmu zmiennej *median_house_price* w modelu. Histogram dla zmiennej zlogarytmowanej przypomina rozkład normalny. Chcąc sprawdzić przypuszczenia, przeprowadzono test na normalność rozkładu Jarque-Bera – zarówno dla *median_house_price* oraz jej zlogarytmowanej postaci odrzucono hipotezę o rozkładzie normalnym. Jednak w przypadku testu dla *ln_median_house_price* zauważono wyższą wartość p-value (0.0002811).

Rys. 5. Histogram dla zmiennej opisującej medianę cen za 1 m² lokali mieszkaniowych w wersji podstawowej (*median_house_price*) oraz zlogarytmowanej (*ln_median_house_price*)

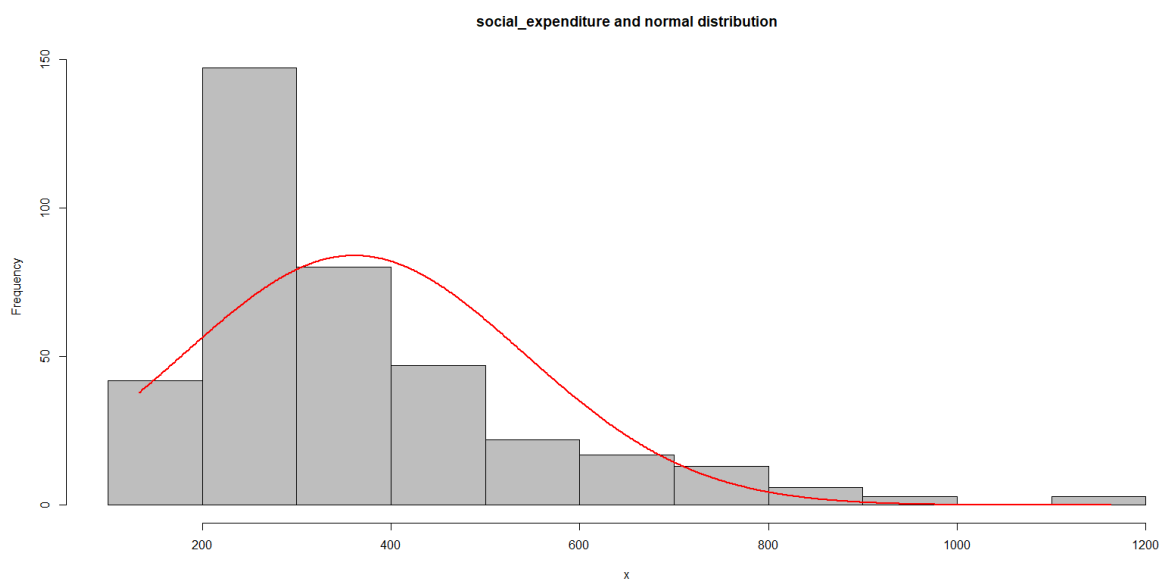


Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

Zmienna opisująca wydatki na pomoc społeczną (*social_expenditure*)

Zmienna opisująca wydatki budżetów powiatów (i miast na prawach powiatu) na pomoc społeczną na jednego mieszkańca nie ma rozkładu normalnego, co zostało zweryfikowane testem Jarque-Bera. Średnio powiat wydaje na pomoc społeczną na jednego mieszkańca 360.90 złotych.

Rys. 6. Histogram zmiennej *social_expenditure*

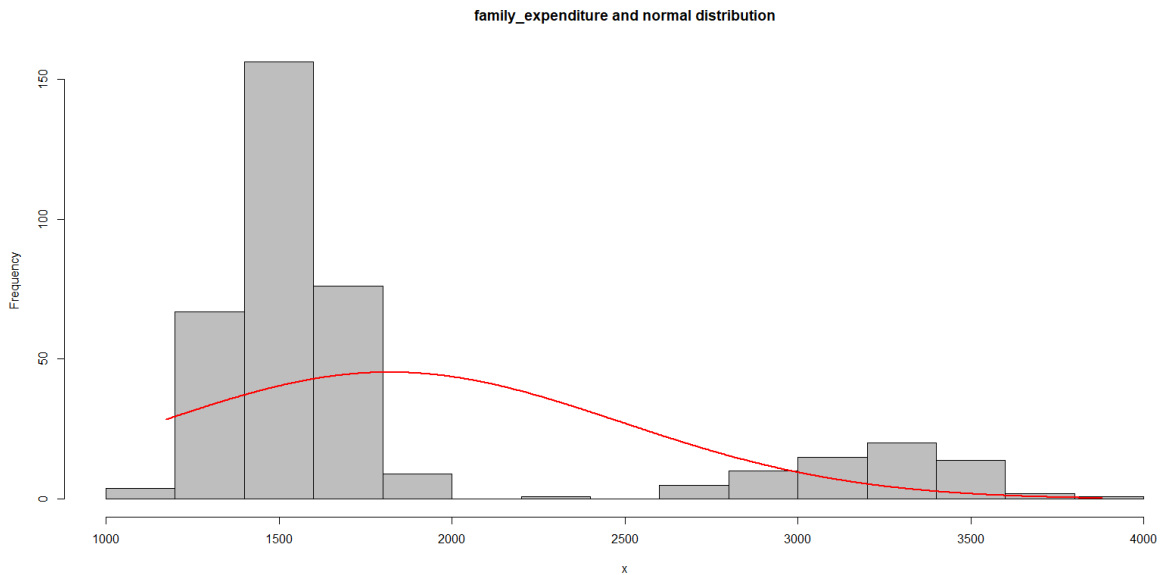


Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

Zmienna opisująca wydatki na rodzinę (*family_expenditure*)

Rozkład zmiennej dotyczącej wydatków budżetów powiatów (lub miast na prawach powiatu) na jednego mieszkańca jest nietypowy. Zdecydowanie najwięcej powiatów wydaje około 1500 złotych na jednego mieszkańca w tym obszarze. Maksymalna wartość zmiennej wynosi 3877.99 złotych.

Rys. 7. Histogram zmiennej *family_expenditure*



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* z wykorzystaniem biblioteki *rcompanion*.

ROZDZIAŁ IV

Formy funkcyjne

W rozdziale 3. punkcie 3.3.1., gdzie analizowano współczynnik urodzeń wskazano, że logarytm ze zmiennej *birth_rate* ma rozkład normalny, dlatego szacuję model ekonometryczny ze zmienną zależną *ln_birth rate*. Przedstawiam kolejne iteracje modelu wraz z uzasadnieniem ich zastosowania.

Postać analityczna wstępnego, pierwszego równania regresji wygląda następująco:

$$\begin{aligned} \ln birth.rate = & \beta_0 + \beta_1 covid19.cases + \beta_2 covid19.deaths + \beta_3 covid19.quarantine \\ & + \beta_4 marriage.rate + \beta_5 divorce.rate + \beta_6 budget.reve.pc \\ & + \beta_7 unemployment.rate + \beta_8 women.reproductive + \beta_9 femininity.ratio \\ & + \beta_{10} avg.salary + \beta_{11} women.working + \beta_{12} men.working \\ & + \beta_{13} median.house.price + \beta_{14} house.ratio + \beta_{15} houses.area.pc \\ & + \beta_{16} avg.people.per.house + \beta_{17} education.expenditure \\ & + \beta_{18} health.expenditure + \beta_{19} social.expenditure + \beta_{20} family.expenditure \\ & + \beta_{21} children + \beta_{22} nursery.places + \beta_{23} doctors + \beta_{24} urbanisation.rate \\ & + \beta_{25} bus.stops + \varepsilon \end{aligned}$$

Przyjmując poziom istotności $\alpha = 5\%$, hipoteza o łącznej nieistotności zmiennych w modelu 1. (Załącznik 2.) zostaje odrzucona ($p\text{-value} < 0.05$). Model tłumaczy w 76.71% zmienność logarytmu współczynnika urodzeń. Jednak 13 zmiennych jest pojedynczo nieistotnych statystycznie ($p\text{-value} > 0.05$), w związku z tym przeprowadzono kolejne iteracje

pozwalające na stworzenie wersji modelu zawierającej jedynie zmienne pojedynczo istotne statystycznie.

W kolejnym kroku dodano zmienne będące potęgami oraz interakcjami wcześniej zastosowanych zmiennych. Do modelu dodano interakcję pomiędzy zmiennymi *nursery_places* a *children* (chcąc przeanalizować zjawisko synergii między liczbą dzieci w wieku do 2 lat oraz liczbą miejsc w żłobkach w danym powiecie lub mieście na prawach powiatu). Kolejna interakcja zmiennych zachodzi pomiędzy *education_expenditure* a *children*, również i w tym przypadku chcę zbadać czy nastąpi wzmocnienie efektu wpływu zmiennych na współczynnik urodzeń.

$$\begin{aligned} \ln birth.rate = & \beta_0 + \beta_1 covid19.cases + \beta_2 covid19.deaths + \beta_3 covid19.quarantine \\ & + \beta_4 marriage.rate + \beta_5 divorce.rate + \beta_6 budget.reve.pc \\ & + \beta_7 unemployment.rate + \beta_8 women.reproductive + \beta_9 femininity.ratio \\ & + \beta_{10} avg.salary + \beta_{11} women.working + \beta_{12} men.working \\ & + \beta_{13} median.house.price + \beta_{14} house.ratio + \beta_{15} houses.area.pc \\ & + \beta_{16} houses.area.pc^2 + \beta_{17} avg.people.per.house \\ & + \beta_{18} education.expenditure + \beta_{19} health.expenditure \\ & + \beta_{20} social.expenditure + \beta_{21} family.expenditure + \beta_{22} children \\ & + \beta_{23} nursery.places + \beta_{24} doctors + \beta_{25} urbanisation.rate + \beta_{26} bus.stops \\ & + \beta_{27} nursery.places * children + \beta_{28} education.expenditure * children + \varepsilon \end{aligned}$$

Wyniki regresji modelu (Załącznik 3.) poprawiły się w stosunku do wstępnej postaci, ponieważ skorygowany współczynnik determinacji wynosi 75.4% (w porównaniu z 75.06%). Iteracja ta tłumaczy 77.22% zmienności zlogarytmowanego współczynnika urodzeń. Pomimo, że wszystkie zmienne są łącznie istotne statystycznie, to 15 z nich jest pojedynczo nieistotna statystycznie, więc wnioskowanie na ich podstawie będzie obarczone błędem.

Chcę również uwzględnić w modelu zlogarytmowane zmienne zaproponowane przez Ermischa w 1998 roku (*ln_unemployment_rate*, *ln_avg_salary* oraz *ln_median_house_price*) dlatego dokonuje kolejnej iteracji.

$$\begin{aligned} \ln birth.rate = & \beta_0 + \beta_1 covid19.cases + \beta_2 covid19.deaths + \beta_3 covid19.quarantine \\ & + \beta_4 marriage.rate + \beta_5 divorce.rate + \beta_6 budget.reve.pc \\ & + \beta_7 \ln unemployment.rate + \beta_8 women.reproductive + \beta_9 femininity.ratio \\ & + \beta_{10} \ln avg.salary + \beta_{11} women.working + \beta_{12} men.working \\ & + \beta_{13} \ln median.house.price + \beta_{14} house.ratio + \beta_{15} houses.area.pc \\ & + \beta_{16} houses.area.pc^2 + \beta_{17} avg.people.per.house \\ & + \beta_{18} education.expenditure + \beta_{19} health.expenditure \\ & + \beta_{20} social.expenditure + \beta_{21} family.expenditure + \beta_{22} children \\ & + \beta_{23} nursery.places + \beta_{24} doctors + \beta_{25} urbanisation.rate + \beta_{26} bus.stops \\ & + \beta_{27} nursery.places * children + \beta_{28} education.expenditure * children + \varepsilon \end{aligned}$$

Rezultaty regresji liniowej dla modelu w wersji przejściowej (Załącznik 4.) poprawiły się o 0.18 punkta procentowego w stosunku do poprzedniej iteracji. Model tłumaczy 77.39%

zmienności logarytmu współczynnika wynagrodzeń, zmienne zastosowane w modelu są łącznie istotne statystycznie, ale nadal wiele z nich jest pojedynczo nieistotna. W związku z tym przeprowadzono procedurę GETS (*General-To-Specific*), czyli metodę „od ogółu do szczegółu” w celu usunięcia zmiennych pojedynczo nieistotnych statystycznie.

Początkowo usunięto zmienne: *covid19_quarantine*, *avg_people_per_house*, interakcję zmiennych *nursery_places* i *children*, zmienną *women_working*, potęgę *houses_area_pc²*, *urbanisation_rate*, *houses_area_pc*, *budget_reve_pc*, *covid19_cases*, *education_expenditure*, *femininity_ratio* oraz *nursery_places*. Test hipotezy złożonej stanowiącej o łącznej nieistotności statystycznej powyższych zmiennych wykazał, że nie ma podstaw aby odrzucić hipotezę zerową (przy założeniu poziomu istotności $\alpha = 0.05$). Pozostałe zmienne nieistotne statystycznie cechowały się p-value oscylującym wokół 10%, dlatego postanowiłam usuwać je pojedynczo.

Wersja ostateczna modelu (Załącznik 4.) tłumaczy zmienność zmiennej zależnej w 75.75%, porównując za pomocą wartości skorygowanego R^2 wersje modeli można uznać tą za najgorszą. Pomimo, że usuwanie zmiennych pogarsza jakość dopasowania, to z drugiej strony zwiększa dokładność oszacowań parametrów dla zmiennych istotnych (Mycielski, 2010).

ROZDZIAŁ V

Diagnostyka modelu

W celu zweryfikowania założeń Klasycznego Modelu Regresji Liniowej (KMRL) przeprowadzono testy diagnostyczne. Sprawdzenie założeń KMRL jest kluczowe, ponieważ dzięki nim możemy opierać się na poprawności estymatorów i móc na ich podstawie tworzyć wnioski (Mycielski, 2010).

5.1. Założenie o prawidłowej formie funkcyjnej

Podczas weryfikacji hipotezy o liniowej formie modelu wykorzystano test RESET (Regression Specification Error Test) w wersji dopasowanej (*fitted*). Testowana hipoteza zerowa oraz hipoteza alternatywa (zakładająca nieliniowość) według Mycielskiego (2010) brzmi następująco:

$$H_0: y = X * \beta + \varepsilon$$

$$H_1: y = f(X * \beta) + \varepsilon$$

Test RESET dla ostatecznej wersji modelu (Załącznik 4.) wskazuje brak podstaw do odrzucenia hipotezy zerowej o poprawności formy funkcyjnej dla przyjętego poziomu istotności $\alpha = 0.05$ (wartość p-value wynosi 0.3377). Założenie o liniowej formie funkcyjnej świadczy, że estymatory parametrów w modelu nie są obciążone, więc ich interpretowanie pozwala na wyciągnięcie statystycznie poprawnych wniosków.

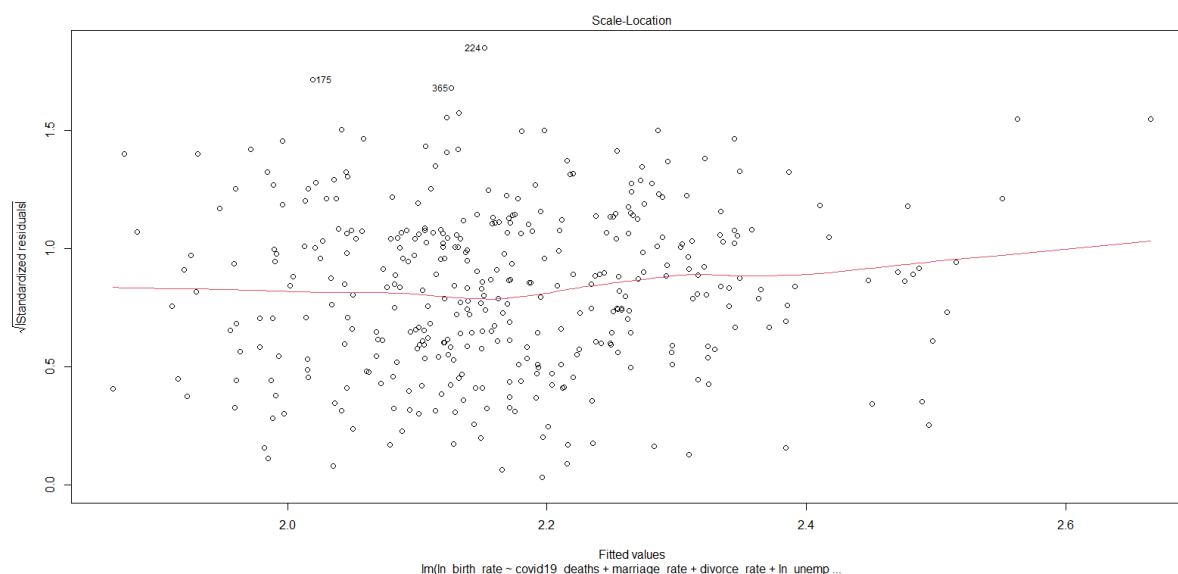
5.2. Założenie o homoskedastyczności reszt ε

Weryfikację homoskedastyczności składnika losowego przeprowadzono za pomocą testu Breuscha-Pagana, którego hipoteza zerowa mówi o braku heteroskedastyczności. Dla przyjętego poziomu istotności 0.05 nie ma podstaw do odrzucenia hipotezy zerowej (p-value = 0.975).

W przypadku niespełnienia założenia o homoskedastyczności reszt Klasycznej Metody Regresji Liniowej mamy do czynienia z brakiem sferyczności błędu losowego, co skutkuje niezgodnością estymatora czego konsekwencją mogą być błędne wyniki wnioskowania statystycznego (Mycielski, 2010). Jednym ze sposobów rozwiązania tego problemu jest zastosowanie estymatora odpornego macierzy wariancji kowariancji White'a, Ważonej Metody Najmniejszych Kwadratów (WMNK) lub Stosowalnej Uogólnionej Metody Najmniejszych Kwadratów (SUMNK). Według rezultatów regresji dla MNK, SUMNK i WMNK (Załącznik 5.) możemy stwierdzić, że niewielkie różnice oszacowań parametrów między zastosowanymi modelami są skutkiem spełnienia założenia o homoskedastyczności reszt.

Dla potwierdzenia testu Breuscha-Pagana skorzystano z analizy wykresów, z których można wnioskować o homogeniczności wariancji. Patrząc na wykres typu *scale-location* (Rys. 7.) można zauważyć, że czerwona linia jest w przybliżeniu pozioma, czyli średnia wielkość standaryzowanych reszt nie zmienia się w funkcji dopasowanych wartości. Natomiast trudno jasno określić czy rozrzut punktów w obrębie czerwonej linii nie zmienia się w zależności od wartości dopasowanych (to znaczy czy wariancja jest homogeniczna).

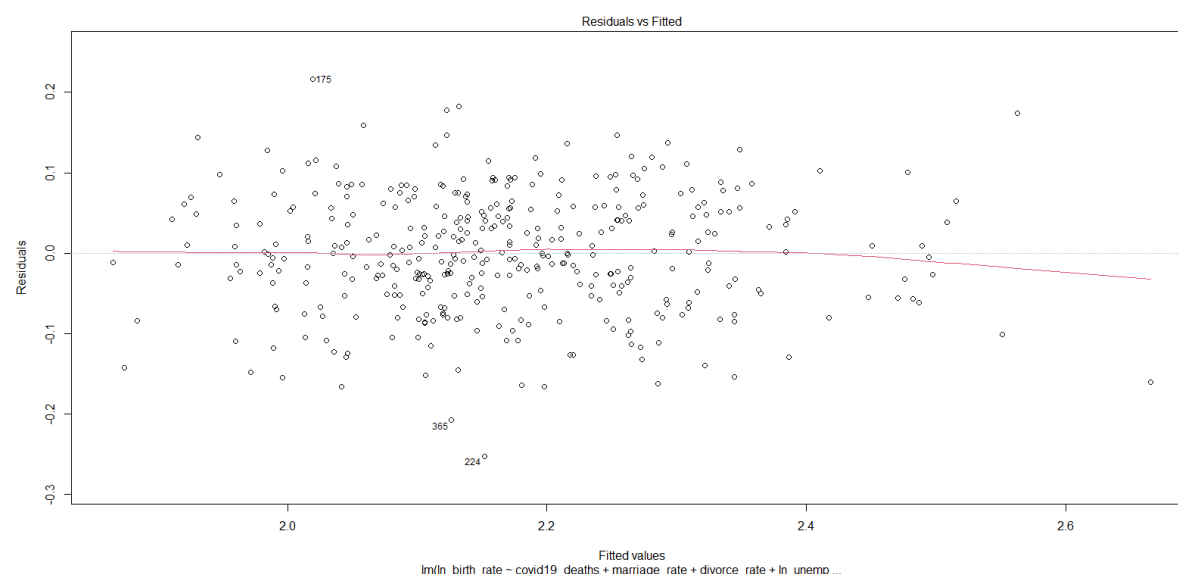
Rys. 7. Wykres *Scale-Location* dla ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia*.

Dla potwierdzenia założenia o homoskedastyczności reszt można również skorzystać z analizy wykresu reszduów i wartości dopasowanych przez model (Rys. 8.). Czerwona linia znajduje się w bliskim położeniu linii przerywanej, natomiast ponownie nie można jasno stwierdzić czy wariancja reszduów jest stała wraz ze wzrostem wartości prognozowanych. Na wykresie zaznaczono obserwacje, które mogą być obserwacjami nietypowymi (ze względu na wysokie wartości reszduów).

Rys. 8. Wykres *Residual vs Fitted* dla ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia*.

5.3. Założenie o normalności rozkładu reszt ε

Żeby sprawdzić czy możliwe jest przeprowadzenie wnioskowania statystycznego za pomocą standardowych metod przeprowadzono testowanie normalności składników losowych za pomocą testu Jarque-Bera. W powyższym teście hipoteza zerowa zakłada normalność rozkładu błędów losowych, natomiast hipoteza alternatywna jego brak (Mycielski, 2010). Testowane hipotezy:

$$H_0: \varepsilon \sim N(0, \sigma^2 I)$$

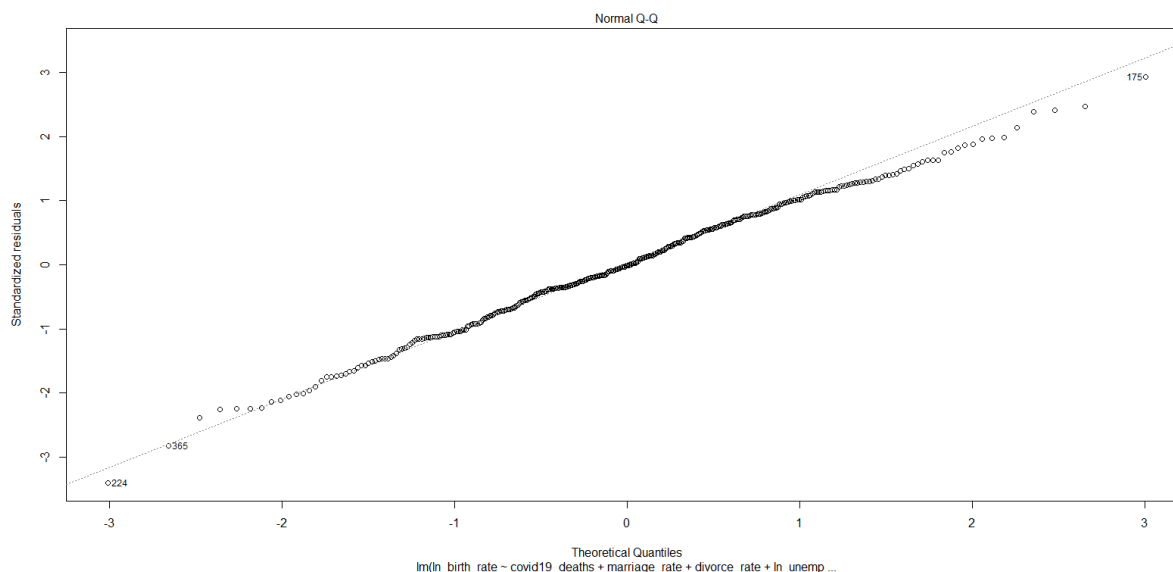
$$H_1: \varepsilon \not\sim N(0, \sigma^2 I)$$

Test Jarque-bera dla modelu w wersji ostatecznej (Załącznik 4.) wskazuje na brak podstaw do odrzucenia hipotezy zerowej o normalności rozkładu reszt na poziomie istotności równym 0.05 (p-value = 0.2998). Założenie o normalności rozkładu reszt jest spełnione, dzięki temu można przeprowadzać prawidłowe wnioskowanie statystyczne w małych próbach.

Testem, który również sprawdza normalność rozkładu zmiennej jest test Shapiro-Wilka. Zweryfikowano hipotezę zerową o rozkładzie normalnych reszt i wykazano brak podstaw do jej odrzucenia (p-value = 0.568 przy założonym poziomie istotności 0.05).

Żeby potwierdzić wyniki testu Jarque-Bera i Shapiro-Wilka przeprowadzono analizę wykresów badających normalność rozkładu składników losowych. Na wykresie kwantylowym (Rys. 9.) możemy zauważyć, że punkty po prawej stronie wykresu leżą poniżej linii przerywanej, co może sugerować brak normalności residuów.

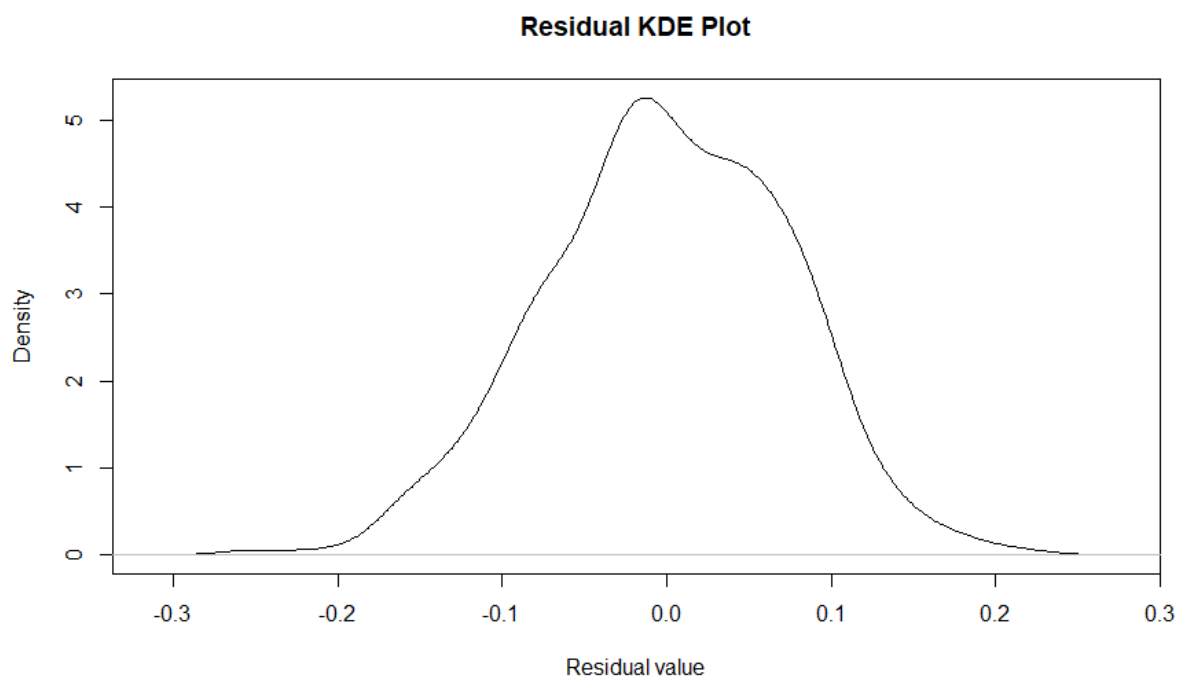
Rys. 9. Wykres kwantylowy dla rozkładu normalnego dla ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia*.

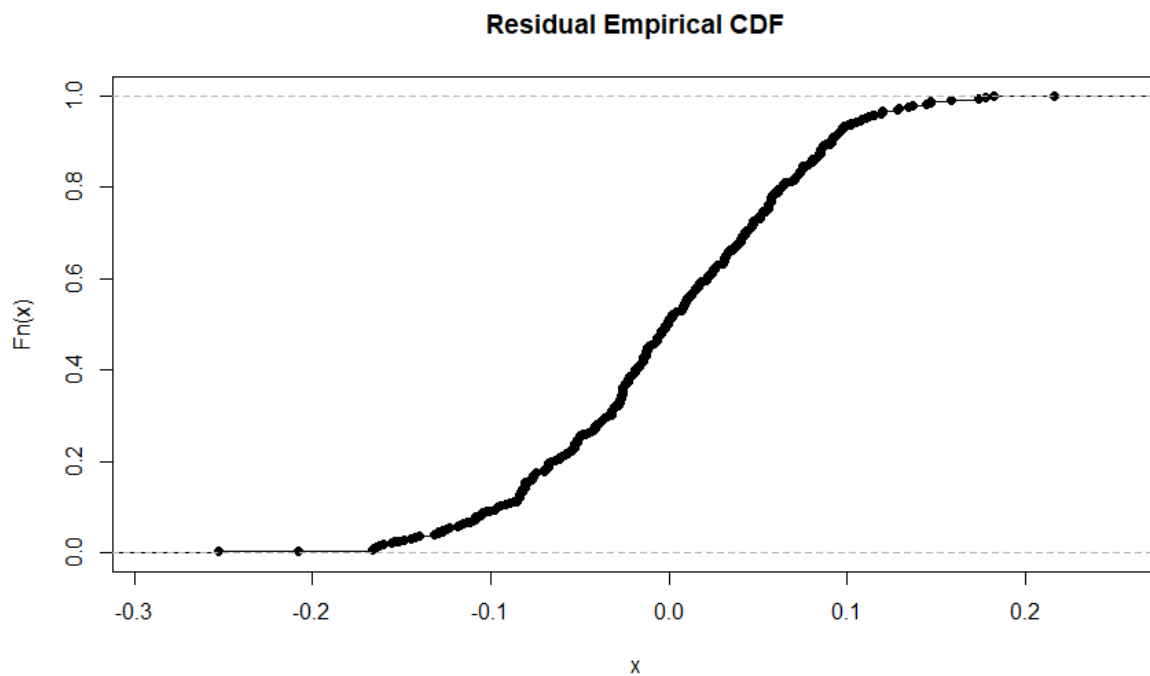
Analizując kwestię normalności rozkładu reszt można wykorzystać także jądrowy estymator gęstości residuów (Rys. 10.). Na podstawie wykresu można wnioskować, że rozkład jest w niewielkim stopniu lewostronnie asymetryczny. Badając normalność rozkładów reszt można też posłużyć się wykresem dystrybuanty rozkładu (Rys. 11.), który w przypadku reszt ostatecznej wersji modelu (Załącznik 4.) przypomina wykres dystrybuanty rozkładu normalnego.

Rys. 10. Jądrowy estymator gęstości dla reszt ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia*.

Rys. 11. Wykres dystrybuanty rozkładu reszt ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia*.

5.4. Założenie o braku autokorelacji czynników losowych ε

W celu zweryfikowania autokorelacji reszt przeprowadzono test Durбина-Watsona oraz Breuscha-Godfrey'a. Oba testy wykazały, że hipotezy zerowe o braku autokorelacji (przy przyjętym poziomie istotności $\alpha = 0.05$) zostały odrzucone (p-value dla testu Durбина-Watsona wynosi 0.000007595, natomiast dla Breuscha-Godfrey'a 0.00003686). Gdy założenie o braku autokorelacji Klasycznej Metody Regresji Liniowej nie jest spełnione mówimy o niesferyczności błędów losowych – charakteryzują się wtedy całkowicie chaotycznymi fluktuacjami (Mycielski, 2010). Jednymi ze sposobów radzenia sobie z tym problemem są: użycie estymatora Newey'a – Westa macierzy wariancji kowariancji lub zastosowanie Stosowalnej Uogólnionej Metody Najmniejszych Kwadratów z estymatorem Prais-Winstena.

Estymator Newey'a-West'a macierzy wariancji kowariancji oraz ostateczna wersja modelu nie różnią się znacząco. Parametry zmiennych nie zmieniły się, jedyne różnice są widoczne w błędach standardowych (Załącznik 6.). Jest to związane z typem danych użytych w pracy – model jest zbudowany na podstawie danych przekrojowych, a problem autokorelacji jest charakterystyczny dla szeregów czasowych.

5.5. Współliniowość

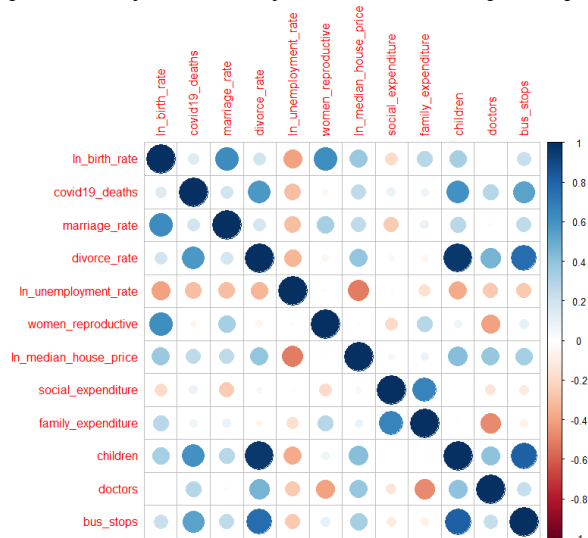
Współliniowość zmiennych objaśniających dotyczy przypadku, w którym występuje silna korelacja między nimi. W przypadku modelowania danych ekonomicznych (między innymi w niniejszej pracy) zjawisko to występuje bardzo często, jest jednak niekorzystne, ponieważ zmniejsza precyzję oszacowań parametrów. Zmierzenie wpływu korelacji na wynik estymacji modelu mierzy statystyka VIF (Variance Inflation Factor) oparta na współczynniku determinacji R^2 (Mycielski, 2010).

Przeprowadzono wyliczenie współczynnika VIF dla ostatecznego modelu (Załącznik 4.). W literaturze podaje się, że gdy $VIF > 10$, to można mówić o występowaniu silnej niedokładnej współliniowości między zmiennymi niezależnymi. W przypadku końcowej wersji modelu zmiennymi, dla których współczynnik przyjmuje wartości większe niż 10 to: *divorce_rate* oraz *children*.

Jedną z metod rozwiązania problemu niedokładnej współliniowości jest usunięcie zmiennej o najwyższym współczynniku VIF (w tej pracy: *children* mierząca liczbę ludności w danym powiecie (lub mieście na prawach powiatu) w wieku do 2 lat, dla której VIF wynosi 25.632833). Jednakże usunięcie tej zmiennej (która jest istotna statystycznie) może

spowodować obciążenie estymatorów przy zmiennych, z którymi jest skorelowana, dlatego zdecydowano o pozostawieniu ich w estymowanym modelu ekonometrycznym.

Rys. 12. Wykres korelacji zmiennych zawartych w ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych Banku Danych Lokalnych oraz Ministerstwa Zdrowia z wykorzystaniem biblioteki *corrplot*.

Z wykresu korelacji zmiennych (Rys. 12.) możemy zauważyć, że zmienne *children* (o najwyższej wartości VIF) jak i *divorce_rate* są skorelowane w znacznym stopniu ze sobą oraz zmienną dotyczącą pandemii COVID-19 i *bus_stops*.

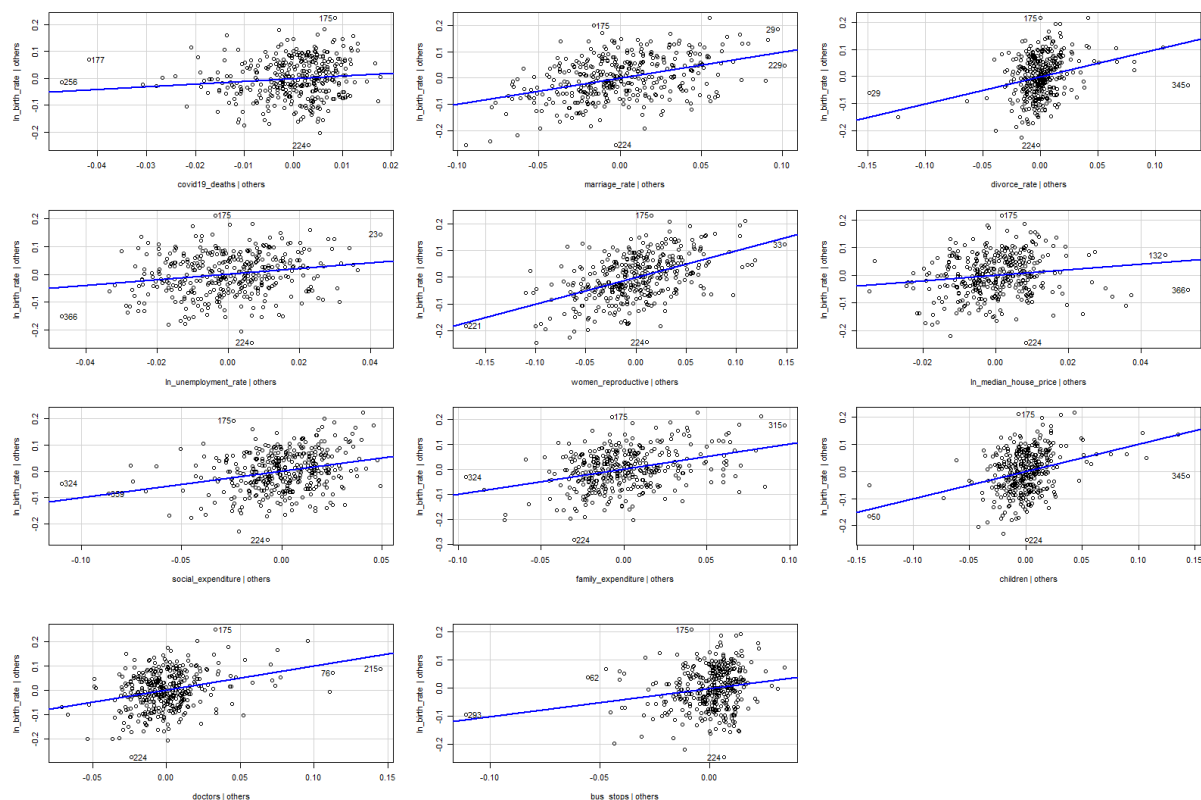
5.6. Obserwacje nietypowe i błędne

Kolejnym etapem analizy jest próba wykrycia zanieczyszczeń bazy danych użytej do estymowania modelu ekonometrycznego. Mogą być związane z wystąpieniami obserwacji nietypowych (nietypowych w stosunku do pozostałych obserwacji) lub błędnych (których powstania nie da się tłumaczyć modelem teoretycznym).

5.6.1. Dźwignia (leverage)

Jednym ze sposobów na wykrycie obserwacji nietypowych jest zastosowanie statystyki nazywanej dźwignią (*leverage*). Nieformalna reguła przyjęta w literaturze mówi, że obserwację powyżej wartości $\frac{2K}{N}$, gdzie K – liczba zmiennych w modelu, N – liczba obserwacji. W przypadku ostatecznej wersji modelu wartością graniczną jest 0,063158.

Rys. 13. Wykresy dźwigni dla ostatecznej wersji modelu (Załącznik 4.) z zaznaczonymi obserwacjami nietypowymi



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia* z wykorzystaniem biblioteki *car*.

Z wykresów dźwigni (Rys. 13.) możemy zaobserwować, że obserwacjami, które najczęściej pojawiają się na wykresach jako obserwacje nietypowe są: 175, 224. Na podstawie porównania podstawowych statystyk zmiennych a wartościami obserwacji można zauważyć, iż na uwagę zasługuje liczba zgonów spowodowanych pandemią koronawirusa. W szczególności obserwacja numer 175, gdzie liczba przypadków osób chorych na COVID-19 osiąga wartość powyżej pierwszego kwartyła, a liczba zgonów wynosi 0. W przypadku obserwacji numer 224 liczba zgonów jest równa zero, jednakże liczba zachorowań w porównaniu do innych powiatów jest niewielka.

5.6.2. Standaryzowane reszty

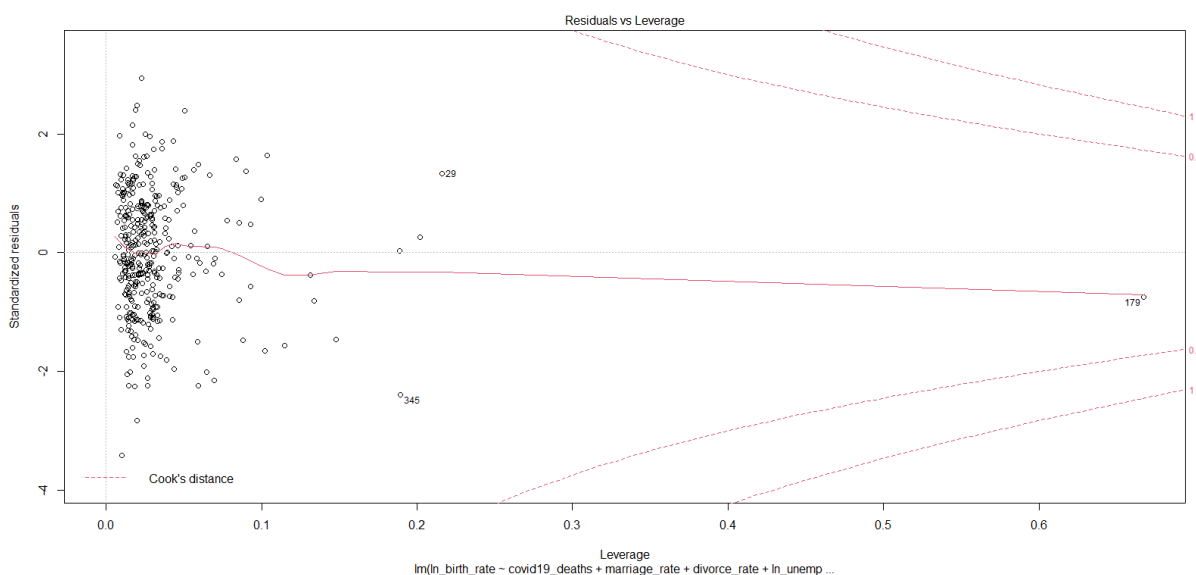
Kolejną statystyką służącą do wykrywania obserwacji nietypowych są reszty standardowe. Diagnozując model testem Jarque-Bera potwierdzono, iż reszty mają rozkład normalny, dlatego wystandaryzowana reszta również ma rozkład normalny. Przyjmuje się,

że obserwację można uznać za nietypową jeśli wartość bezwzględna wystandaryzowanej reszty będzie przyjmować wartości większe od 2 (Mycielski, 2010).

Dla ostatecznej wersji modelu (Załącznik 4.) odnotowano 17 obserwacji, których wartość bezwzględna jest wyższa od 2. Największe wartości odnotowały obserwacje numer: 224 (3,408177) oraz 175 (2,936103).

Na wykresie (Rys. 14.) możemy sprawdzić jak zmienia się rozrzut standaryzowanych reszt wraz ze wzrostem dźwigni, ale także zaobserwować obserwacje wpływowe (29, 179, 345), których usunięcie mogłoby spowodować znaczne zmiany parametrów modelu.

Rys. 14. Wykres standaryzowanych reszt i dźwigni dla ostatecznej wersji modelu (Załącznik 4.)

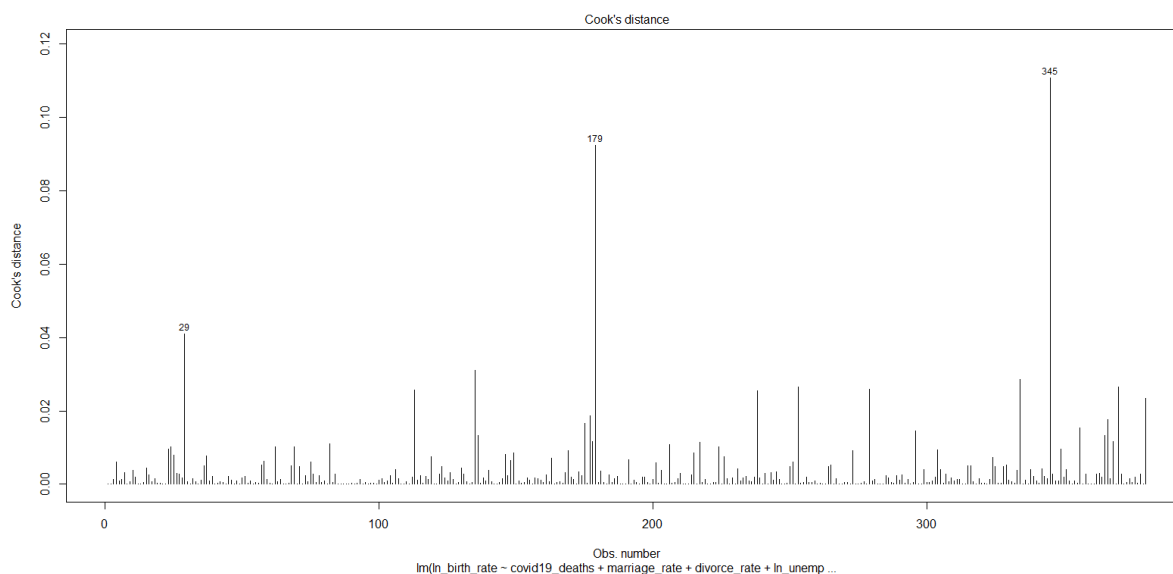


Źródło: Opracowanie własne na podstawie danych Banku Danych Lokalnych oraz Ministerstwa Zdrowia.

5.6.3. Odległość Cooka

Statystyka Cooka mierzy wpływ pojedynczej obserwacji na wynik regresji liniowej. Nieformalna reguła mówi, że powinniśmy przyjrzeć się obserwacjom, dla których odległość Cooka jest wyższa od $\frac{4}{N}$ (w tym modelu gdy jest większa niż 0,010526). Na podstawie wykresu (Rys. 15.) możemy zauważyć trzy obserwacje nietypowe: 29, 179, 345.

Rys. 15. Wykres przedstawiający odległości Cooka dla ostatecznej wersji modelu (Załącznik 4.)



Źródło: Opracowanie własne na podstawie danych *Banku Danych Lokalnych* oraz *Ministerstwa Zdrowia*.

Na wykresie pojawiły się wartości, które zanotowano w przypadku standaryzowanych reszt i dźwigni, dlatego postanowiono stworzyć model wykluczając obserwacje, które mogą być outlierami, to jest 29, 175 i 345. W wyniku ich usunięcia możemy zaobserwować niewielkie zmiany wartości parametrów β zmiennych objaśniających, a także spadek wartości współczynnika determinacji i jego skorygowanej wersji (Załącznik 7.). Mimo wszystko zdecydowałam, aby wnioskowanie statystyczne prowadzić na wersji modelu ekonometrycznego zbudowanej na pełnej bazie danych zawierającej 380 obserwacji.

ROZDZIAŁ VI

Weryfikacja hipotez

6.1. Interpretacja wyników estymacji

W wyniku przeprowadzenia kolejnych iteracji modelu, finalna postać funkcyjna prezentuje się następująco:

$$\ln birth.rate = 0.0498 - 0.0004 covid19.deaths + 0.0929 marriage.rate - 0.4357 divorce.rate - 0.0372 \ln unemployment.rate + 3.0904 women.working +$$

$$0.0437 \ln median.house.price - 0.0002 social.expenditure + 0.0001 family.expenditure + 0.00002 children + 0.0011 doctors - 0.0001 bus.stops$$

Na podstawie wartości współczynnika determinacji wiadomo, że model wyjaśnia zmienność logarytmu współczynnika urodzeń w 75.75%. Na poziomie istotności $\alpha = 0.05$ wszystkie zmienne są łącznie oraz pojedynczo istotne.

Na podstawie oszacowań parametrów możemy wyciągnąć następujące wnioski:

- śmierć na skutek COVID-19 spowoduje zmniejszenie współczynnika urodzeń o 0.04%, ceteris paribus;
- zwiększenie współczynnika małżeństw o jednostkę spowoduje wzrost współczynnika urodzeń o 9.73%, ceteris paribus;
- zwiększenie współczynnika rozwodów o jednostkę spowoduje spadek współczynnika urodzeń o 35.32%, ceteris paribus;
- wzrost stopy bezrobocia o 1% spowoduje spadek współczynnika urodzeń o 3.72%, ceteris paribus;
- zwiększenie odsetka kobiet w wieku reprodukcyjnym o jednostkę sprawi, że współczynnik urodzeń zwiększy się o 2098.6%, ceteris paribus;
- wzrost mediany cen na rynku mieszkaniowym o 1% spowoduje zwiększenie współczynnika urodzeń o 4,37%, ceteris paribus;
- wzrost wydatków na pomoc społeczną o jednostkę sprawi, że zmienna objaśniana zmniejszy się o 0.021%, ceteris paribus;
- w przypadku zwiększenia wydatków na rodzinę o jednostkę, zmienna zależna wzrośnie o 0.008%, ceteris paribus;
- dodatkowa osoba w grupie wiekowej do 2 lat spowoduje zwiększenie współczynnika urodzeń o 0.002%, ceteris paribus;
- wzrost odsetka personelu medycznego na 10 tysięcy mieszkańców o jednostkę wpłynie na wzrost współczynnika urodzeń o 0.114%, ceteris paribus;
- dodatkowy przystanek autobusowy lub tramwajowy sprawi, że współczynnik urodzeń zmaleje o 0.007%, ceteris paribus.

6.2. Weryfikacja hipotez badawczych

H0: Pandemia COVID-19 negatywnie wpływa na liczbę urodzonych dzieci – ze względu na wzrost bezrobocia spowodowany zamrożeniem gospodarki.

Estymacje parametrów modelu potwierdziły założenie, że wzrost bezrobocia ujemnie wpływa na współczynnik urodzeń. W porównaniu z przewidywaniami Kearney i Levine’a (2020) negatywny efekt jest mniejszy, ponieważ model przewiduje, że wzrost stopy bezrobocia o 1% spowoduje spadek wskaźnika o 3.72%.

H1: Pandemia COVID-19 negatywnie wpływa na liczbę urodzonych dzieci – ze względu na śmiertelność.

Hipoteza została potwierdzona, zgony spowodowane chorobą wywołaną SARS-Cov-2 wpływają ujemnie na współczynnik urodzeń. Ryzyko śmierci jest wysokie wśród osób starszych, dlatego można wnioskować, że potencjalni rodzice nie decydują się mieć potomstwa ze względu na brak instytucji dziadków. Jednak warto wspomnieć, że zgony występowały również wśród osób młodych (w wieku reprodukcyjnym). Z parametru przy zmiennej dotyczącej kobiet w wieku reprodukcyjnym można snuć wnioski o tym, iż zgony wśród grupy ludzi w wieku 15-49 lat wpływają na współczynnik urodzeń jeszcze bardziej.

H2: Zwiększenie wydatków na dzieci powoduje wzrost współczynnika urodzeń.

Chcąc zbadać postawioną hipotezę, należy odpowiedzieć na pytanie jak na współczynnik urodzeń wpłyną dodatkowe jednostki wydatków budżetów powiatów na pomoc społeczną i rodzinę.

$$\begin{aligned} E(\ln birth.rate | social.expenditure = x + 1, family.expenditure = y + 1) \\ - E(\ln birth.rate | social.expenditure = x, family.expenditure = y) \\ = -0.0002(x + 1) + 0.0001(y + 1) + 0.0002x - 0.0001y = -0.0001 \end{aligned}$$

Paradoksalnie dodatkowe jednostki pieniężne przeznaczone pośrednio na dzieci powodują spadek wzrostu współczynnika urodzeń o 0.1%. Jest to sprzeczne między innymi z wynikami badań Kalwija (2010), który potwierdził empirycznie dodatnią zależność między wdrażaniem narzędzi polityki prorodzinnej przez państwa a wzrostem dzietności. Odrzucenie powyższej hipotezy może być spowodowane tym, że zaproponowane zmienne mogą oddziaływać na podniesienie jakości dzieci jako dobra, a nie ich liczby. Powiaty w ramach budżetu przeznaczonego na pomoc społeczną mogą przekazywać pieniądze rodzinom potrzebującym, które poprawią dobrostan dziecka, a nie przeznaczają je na wydanie na świat nowego potomka.

H3: Wzrost cen mieszkań na rynku ujemnie koreluje ze współczynnikiem urodzeń.

Hipoteza ta została odrzucona, ponieważ z modelu wynika, że wzrost mediany cen za metr kwadratowy lokalu mieszkaniowego o 1% spowoduje wzrost współczynnika urodzeń o 4.37%. Taką zależność zasygnalizowało kilkoro badaczy, na przykład Dettling i Kearney (2014), które wykazały, że wzrost cen domów o 10 000 dolarów przełoży się na wzrost urodzeń o 0.8%. Wskazują również, że zwiększenie narodzin dzieci zauważamy w przypadku właścicieli domów (w przeciwieństwie do spadku rozrodczości wśród osób nieposiadających nieruchomości). Jeśli przyjąłbyśmy założenie o braku różnic między rynkiem mieszkaniowym w Stanach Zjednoczonych a w Polsce, to można wnioskować, że Polacy są w większości właścicielami domów, w których mieszkają.

ZAKOŃCZENIE

Model opracowany w powyższej pracy potwierdził część hipotez. Okazuje się, że większość wyników nie odbiega znacznie od tych przedstawionych w literaturze. Jednak należy ostrożnie wnioskować z oszacowań parametrów modelu, gdyż z powodu braku danych na poziomie powiatów wiele zmiennych mających wpływ na współczynnik urodzeń (potencjalnie istotnych) nie zostało uwzględnionych, co mogło doprowadzić do obciążenia szacunków.

Problem niskiego współczynnika urodzeń jest istotny i powinien być dogłębnie opracowany, ponieważ skutki starzenia się społeczeństwa będziemy odczuwać coraz bardziej. Zbadanie wpływu pandemii na wskaźnik może pomóc stworzyć nowe programy, które będą pomagały podjąć decyzję o urodzeniu dziecka – na przykład tworzeniu nowych miejsc opieki dziennej, aby rodzice mogli łączyć wychowywanie dziecka z pracą.

BIBLIOGRAFIA

- Arcuri, L. (2019). 4 Contributing Factors to Declining Fertility Rates. Empowered Women's Health. Retrieved January 16, 2022, from <https://www.volusonclub.net/empowered-womens-health/4-contributing-factors-to-declining-fertility-rates-a-global-overview/>
- Becker, G. S. (1960). An Economic Analysis of Fertility. National Bureau of Economic Research, 209–240. <https://www.nber.org/system/files/chapters/c2387/c2387.pdf>
- Birth rate, crude (per 1,000 people). (1960–2000). [Baza danych]. World Bank. <https://data.worldbank.org/indicator/SP.DYN.CBRT.IN>
- Black, D. A., Kolesnikova, N., Sanders, S. G., & Taylor, L. J. (2013). Are Children “Normal”? Review of Economics and Statistics, 95(1), 21–33. https://doi.org/10.1162/rest_a_00257
- Chromińska, M. (1998). Czynniki ekonomiczne we współczesnych koncepcjach ludnościowych. Ruch Prawniczy, Ekonomiczny i Socjologiczny, 3, 361–372. https://repozytorium.amu.edu.pl/bitstream/10593/5490/1/24_Maria_Chrominska_Czynnik%20ekonomiczny%20we%20wsp%C3%B3%C5%82czesnych%20koncepcjach_361-372.pdf
- Currie, J., & Schwandt, H. (2014). Short- and long-term effects of unemployment on fertility. Proceedings of the National Academy of Sciences, 111(41), 14734–14739. <https://doi.org/10.1073/pnas.1408975111>
- Dettling, L. J., & Kearney, M. S. (2014). House prices and birth rates: The impact of the real estate market on the decision to have a baby. Journal of Public Economics, 110, 82–100. <https://doi.org/10.1016/j.jpubeco.2013.09.009>
- Ermisch, J. (1988). Economic influences on birth rates. National Institute Economic Review, 126(1), 71–92. <https://doi.org/10.1177/002795018812600108>
- Hryniewicz, J., Witkowski, J., & Potrykowska, A. (2018). Sytuacja demograficzna Polski jako wyzwanie dla polityki społecznej i gospodarczej. Rządowa Rada Ludnościowa.
- Kalwij, A. (2010). The impact of family policy expenditure on fertility in western Europe. Demography, 47(2), 503–519. <https://doi.org/10.1353/dem.0.0104>
- Kearney, M. S., & Wilson, R. (2018). Male Earnings, Marriageable Men, and Nonmarital Fertility: Evidence from the Fracking Boom. The Review of Economics and Statistics, 100(4), 678–690. https://doi.org/10.1162/rest_a_00739
- Lindberg, L. D., VandeVusse, A., Mueller, J., & Kirstein, M. (2020). Early Impacts of the COVID-19 Pandemic: Findings from the 2020 Guttmacher Survey of Reproductive Health Experiences. Guttmacher Institute. Retrieved January 18, 2022, from <https://www.guttmacher.org/report/early-impacts-covid-19-pandemic-findings-2020-guttmacher-survey-reproductive-health>
- Lindo, J. M. (2010). Are Children Really Inferior Goods? Journal of Human Resources, 45(2), 301–327. <https://doi.org/10.3368/jhr.45.2.301>

- Lovenheim, M. F., & Mumford, K. J. (2013). Do Family Wealth Shocks Affect Fertility Choices? Evidence from the Housing Market. *Review of Economics and Statistics*, 95(2h), 464–475. https://doi.org/10.1162/rest_a_00266
- Ministerstwo Zdrowia. (2020). Raport zakażeń koronawirusem (SARS-CoV-2) [Baza danych]. <https://www.gov.pl/web/koronawirus/wykaz-zarazen-koronawirusem-sars-cov-2>
- Mycielski, J. (2010). *Ekonometria*. Wydział Nauk Ekonomicznych, Uniwersytet Warszawski.
- Rogalski, M. (2020). COVID-19 w powiatach [Baza danych]. https://docs.google.com/spreadsheets/d/1Tv6jKMUYdK6ws6SxxAsHVxZbglZfisC8x_HZ1jacmBM/edit#gid=42091932
- Udry, J. R. (1970). The effect of the Great Blackout of 1965 on births in New York City. *Demography*, 7(3), 325–327. <https://doi.org/10.2307/2060151>
- Vandenbroucke, G. (2016, December 13). The Link between Fertility and Income. Federal Reserve Bank of St. Louis. Retrieved January 15, 2022, from <https://www.stlouisfed.org/on-the-economy/2016/december/link-fertility-income>

ZAŁĄCZNIKI

Załącznik 1. Podstawowe statystyki zmiennych ostatecznej wersji modelu

STATISTIC	N	MEAN	ST. DEV.	MIN	PCTL(25)	PCTL(75)	MAX
LN_BIRTH_RATE	380	2.169	0.149	1.732	2.067	2.254	2.736
COVID19_DEATHS	380	23.379	27.376	0	2.8	34	243
MARRIAGE_RATE	380	3.661	0.432	2.600	3.300	4.000	5.100
DIVORCE_RATE	380	0.133	0.212	0.017	0.061	0.142	3.267
LN_UNEMPLOYMENT_RATE	380	1.984	0.490	0.531	1.668	2.354	3.165
WOMEN_REPRODUCTIVE	380	0.449	0.019	0.381	0.437	0.461	0.501
LN_MEDIAN_HOUSE_PRICE	380	8.187	0.319	7.184	7.978	8.375	9.391
SOCIAL_EXPENDITURE	380	360.896	180.647	133.097	237.094	434.694	1,162.852
FAMILY_EXPENDITURE	380	1,819.767	669.389	1,173.412	1,436.539	1,723.992	3,877.993
CHILDREN	380	5,005.884	6,918.297	779	2,410.2	5,296.8	105,383
DOCTORS	380	37.059	32.608	1.000	17.150	45.100	199.400
BUS_STOPS	380	437.532	356.981	82	235.5	540.2	4,342

Załącznik 2. Rezultaty regresji liniowej dla wersji modelu nr 1 (wersja wstępna)

Dependent variable:			
ln_birth_rate			
covid19_cases	0.00001 (0.00000)	avg_people_per_house	-0.014 (0.027)
covid19_deaths	-0.001*** (0.0002)	education_expenditure	0.00000 (0.00002)
covid19_quarantine	0.00000 (0.00000)	health_expenditure	-0.0001* (0.00004)
marriage_rate	0.091*** (0.011)	social_expenditure	-0.0002*** (0.00005)
divorce_rate	-0.410*** (0.093)	family_expenditure	0.0001*** (0.00002)
budget_reve_pc	0.00000 (0.00000)	children	0.00002*** (0.00000)
unemployment_rate	-0.004*** (0.001)	nursery_places	-0.00003** (0.00001)
women_reproductive	3.046*** (0.407)	doctors	0.001*** (0.0002)
femininity_ratio	-0.002 (0.002)	urbanisation_rate	0.0002 (0.0002)
avg_salary	-0.00002* (0.00001)	bus_stops	-0.0001*** (0.00002)
women_working	-0.076 (0.139)	Constant	0.701** (0.351)
men_working	0.196* (0.100)	-----	
median_house_price	0.00001** (0.00000)	Observations	380
house_ratio	-0.0002 (0.001)	R2	0.767
houses_area_pc	-0.001 (0.002)	Adjusted R2	0.751
		Residual Std. Error	0.074
		F Statistic	46.634***
		=====	
		Note:	*p<0.1; **p<0.05; ***p<0.01

Załącznik 3. Rezultaty regresji liniowej dla wersji przejściowej modelu (dodanie potęg i interakcji)

Dependent variable:			
ln_birth_rate			
covid19_cases	0.00000 (0.00000)	houses_area_pc_2	0.0002 (0.0002)
covid19_deaths	-0.001*** (0.0002)	avg_people_per_house	-0.008 (0.027)
covid19_quarantine	-0.000 (0.00000)	education_expenditure	0.00002 (0.00002)
marriage_rate	0.083*** (0.012)	health_expenditure	-0.0001* (0.00004)
divorce_rate	-0.426*** (0.096)	social_expenditure	-0.0002*** (0.00005)
budget_reve_pc	0.00000 (0.00000)	family_expenditure	0.0001*** (0.00002)
unemployment_rate	-0.004*** (0.001)	children	0.00003*** (0.00001)
women_reproductive	2.642*** (0.430)	nursery_places	-0.00002 (0.00001)
femininity_ratio	-0.004 (0.002)	doctors	0.001*** (0.0003)
avg_salary	-0.00002** (0.00001)	urbanisation_rate	0.0001 (0.0002)
women_working	-0.076 (0.138)	bus_stops	-0.0001*** (0.00002)
men_working	0.201** (0.100)	nurseryxchildren	-0.000 (0.000)
median_house_price	0.00001** (0.00000)	education_expenditurexchildren	-0.000* (0.000)
house_ratio	0.003 (0.002)	Constant	1.182*** (0.436)
houses_area_pc	-0.014 (0.015)		
		Observations	380
		R2	0.772
		Adjusted R2	0.754
		Residual Std. Error	0.074
		F Statistic	42.487***
		Note: *p<0.1; **p<0.05; ***p<0.01	

Załącznik 4. Rezultaty regresji liniowej dla wersji przejściowej modelu (zlogarytmowane zmienne) oraz finalnej

Dependent variable:					
	ln_birth_rate				
	wersja przejściowa	wersja finalna			
	(1)	(2)			
covid19_cases	0.00000 (0.00000)		health_expenditure	-0.0001* (0.00004)	
covid19_deaths	-0.001*** (0.0002)	-0.0004** (0.0002)	social_expenditure	-0.0002*** (0.00005)	-0.0002*** (0.00004)
covid19_quarantine	-0.000 (0.00000)		family_expenditure	0.0001*** (0.00002)	0.0001*** (0.00001)
marriage_rate	0.082*** (0.012)	0.093*** (0.011)	children	0.00003*** (0.00001)	0.00002*** (0.00000)
divorce_rate	-0.430*** (0.095)	-0.436*** (0.081)	nursery_places	-0.00002 (0.00001)	
budget_reve_pc	0.00000 (0.00000)		doctors	0.001*** (0.0003)	0.001*** (0.0002)
ln_unemployment_rate	-0.032** (0.013)	-0.037*** (0.010)	urbanisation_rate	0.0001 (0.0002)	
women_reproductive	2.644*** (0.426)	3.090*** (0.284)	bus_stops	-0.0001*** (0.00002)	-0.0001*** (0.00002)
femininity_ratio	-0.004 (0.002)		nurseryxchildren	-0.000 (0.000)	
ln_avg_salary	-0.092* (0.049)		education_expenditurexchildren	-0.000 (0.000)	
women_working	-0.078 (0.137)		Constant	1.532** (0.618)	0.050 (0.173)
men_working	0.182* (0.101)		Observations	380	380
ln_median_house_price	0.049*** (0.018)	0.044*** (0.016)	R2	0.774	0.758
house_ratio	0.002 (0.002)		Adjusted R2	0.756	0.750
houses_area_pc	-0.012 (0.015)		Residual Std. Error	0.074 (df = 351)	0.075 (df = 368)
houses_area_pc_2	0.0002 (0.0002)		F Statistic	42.901*** (df = 28; 351)	104.519*** (df = 11; 368)
avg_people_per_house	-0.008 (0.027)		Note: *p<0.1; **p<0.05; ***p<0.01		
education_expenditure	0.00002 (0.00002)				

Załącznik 5. Rezultaty regresji dla MNK, SUMNK i WMNK

Dependent variable:			
	ln_birth_rate		
	OLS MNK (1)	normal SUMNK (2)	OLS WMNK (3)
covid19_deaths	-0.0004** (0.0002)	-0.0004** (0.0002)	-0.0004** (0.0002)
marriage_rate	0.093*** (0.011)	0.093*** (0.011)	0.093*** (0.011)
divorce_rate	-0.436*** (0.081)	-0.436*** (0.081)	-0.450*** (0.085)
ln_unemployment_rate	-0.037*** (0.010)	-0.037*** (0.010)	-0.037*** (0.010)
women_reproductive	3.090*** (0.284)	3.090*** (0.284)	3.113*** (0.283)
ln_median_house_price	0.044*** (0.016)	0.044*** (0.016)	0.041*** (0.015)
social_expenditure	-0.0002*** (0.00004)	-0.0002*** (0.00004)	-0.0002*** (0.00004)
family_expenditure	0.0001*** (0.00001)	0.0001*** (0.00001)	0.0001*** (0.00001)
children	0.00002*** (0.00000)	0.00002*** (0.00000)	0.00002*** (0.00000)
doctors	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)
bus_stops	-0.0001*** (0.00002)	-0.0001*** (0.00002)	-0.0001*** (0.00002)
Constant	0.050 (0.173)	0.050 (0.173)	0.059 (0.173)
Observations	380	380	380
R2	0.758		0.752
Adjusted R2	0.750		0.745
Log Likelihood		452.643	
Akaike Inf. Crit.		-881.285	
Residual Std. Error (df = 368)	0.075		1.264
F Statistic (df = 11; 368)	104.519***		101.551***
Note: *p<0.1; **p<0.05; ***p<0.01			

Załącznik 6. Estymator Newey'a-West a macierzy wariancji kowariancji

Dependent variable:		
	ln_birth_rate	
	OLS	coefficient
		test
	wersja finalna estymator Neweya-West a	
covid19_deaths	-0.0004** (0.0002)	-0.0004** (0.0002)
marriage_rate	0.093*** (0.011)	0.093*** (0.012)
divorce_rate	-0.436*** (0.081)	-0.436*** (0.089)
ln_unemployment_rate	-0.037*** (0.010)	-0.037*** (0.009)
women_reproductive	3.090*** (0.284)	3.090*** (0.325)
ln_median_house_price	0.044*** (0.016)	0.044** (0.017)
social_expenditure	-0.0002*** (0.00004)	-0.0002*** (0.00004)
family_expenditure	0.0001*** (0.00001)	0.0001*** (0.00001)
children	0.00002*** (0.00000)	0.00002*** (0.00000)
doctors	0.001*** (0.0002)	0.001*** (0.0002)
bus_stops	-0.0001*** (0.00002)	-0.0001*** (0.00002)
Constant	0.050 (0.173)	0.050 (0.215)
Observations	380	
R2	0.758	
Adjusted R2	0.750	
Residual Std. Error	0.075	
F Statistic	104.519***	

Note: *p<0.1; **p<0.05; ***p<0.01

Załącznik 7. Rezultaty regresji finalnej wersji modelu oraz modelu z wykluczeniem obserwacji nietypowych

Dependent variable:		
	ln_birth_rate	
	wersja finalna	wersja bez outlierów
covid19_deaths	-0.0004** (0.0002)	-0.0004** (0.0002)
marriage_rate	0.093*** (0.011)	0.086*** (0.011)
divorce_rate	-0.436*** (0.081)	-0.489*** (0.087)
ln_unemployment_rate	-0.037*** (0.010)	-0.038*** (0.010)
women_reproductive	3.090*** (0.284)	2.982*** (0.286)
ln_median_house_price	0.044*** (0.016)	0.038** (0.016)
social_expenditure	-0.0002*** (0.00004)	-0.0002*** (0.00004)
family_expenditure	0.0001*** (0.00001)	0.0001*** (0.00001)
children	0.00002*** (0.00000)	0.00002*** (0.00000)
doctors	0.001*** (0.0002)	0.001*** (0.0002)
bus_stops	-0.0001*** (0.00002)	-0.0001*** (0.00002)
Constant	0.050 (0.173)	0.161 (0.178)
Observations	380	377
R2	0.758	0.756
Adjusted R2	0.750	0.749
Residual Std. Error	0.075 (df = 368)	0.074 (df = 365)
F Statistic	104.519*** (df = 11; 368)	102.930*** (df = 11; 365)
Note:		
*p<0.1; **p<0.05; ***p<0.01		