ELŻBIETA GOŁATA

ANALIZA ŚCIEŻKOWA W BADANIU UWARUNKOWAŃ DZIETNOŚCI KOBIET

1. Geneza i zastosowania analizy ścieżkowej

Pojęciem przyczyny badacze posługują się bardzo często, rzadko natomiast podają jej definicję bądź choćby wyjaśnienia. Zamiast definiowania przyczyny podaje się zazwyczaj warunki niezbędne dla zachodzenia relacji „*X* jest przyczyną *Y*". Wśród tych warunków wymienia się często bezwyjątkowe następstwo oraz relację czasową, tzn. jeżeli *X* ma być przyczyną *Y*, to skutek (*Y*) powinien nastąpić później aniżeli przyczyna (*X*). Innym dość powszechnie przyjmowanym elementem „definicji" przyczyny jest jej charakter sprawczy. Według H. M. Blalocka [1964] związek przyczynowy to coś więcej aniżeli następstwo zjawisk. Uważa on, że jeżeli *X* jest przyczyną *Y*, to zmiany *X* prowadzą do zmian *Y*, a nie tylko, że po zmianach *X* następują zmiany *Y*.

Jak zauważa J. Karpiński [1985], analizy przyczynowe polegają często na interpretacji zależności statystycznych, które traktuje się jako przyczynowe, jeżeli zachodzą określone relacje czasowe, a nie są one pozorne. Czasem uważa się natomiast, że zależności statystyczne pozwalają jedynie wnioskować o zależności przyczynowej.

Statystyczne metody badania współzależności są coraz bardziej wyrafinowane i dają dobre rezultaty tak w przypadku cech jakościowych, jak i ilościowych. Niemniej jednak, żadna analiza formalna nie udzieli nam odpowiedzi na pytanie o przyczynę i skutek. Również przedstawiona niżej technika analizy ścieżkowej, proponując pewne nowe rozwiązania, może być co najwyżej pomocna w kwestiach interpretacji. Wyraźnie stawia ona wymóg przyczynowego uporządkowania zmiennych *a priori*, na podstawie wcześniejszej analizy merytorycznej.

Metoda analizy ścieżkowej przedstawiona została przez genetyka Sewalla Wrighta [1934, 1960]. Najważniejsze aplikacje tej metody związane były

z badaniem zagadnień genetycznych, gdzie stosowana była z dużym powodzeniem. O. D. Duncan [1966] omawia również możliwości wykorzystania analizy ścieżkowej w socjologii. Pierwsze przykłady zastosowań analizy ścieżkowej w demografii wiążą się z badaniem płodności przeprowadzonym w ramach World Fertility Survey na Fidżi. W swojej pracy M. G. Kendall oraz C. A. O'Muircheartaigh [1977] nie tylko prezentują zasady analizy ścieżkowej, ale także jej różnorodne adaptacje.

Metoda analizy ścieżkowej została przez autorkę szczegółowo omówiona w innym opracowaniu [1992]. Celem niniejszego artykułu jest natomiast jej praktyczne zastosowanie w badaniu zależności między zjawiskami demograficznymi oraz próba porównania z rezultatami otrzymanymi za pomocą analizy korelacji i regresji.

M. G. Kendall oraz C. A. O'Muircheartaigh [1977] omawiają także inne warianty modeli ścieżkowych, a mianowicie tzw. modele „nienasycone" oraz uwzględniające zmienne jakościowe. Model nienasycony to taki, w którym nie wszystkie zmienne są zależne od zmiennych wcześniejszych w łańcuchu uporządkowania przyczynowego. Przykładem może być model zawierający zmienną etniczną określającą rasę. Zmienną taką można przedstawić w postaci zmiennej binarnej; ponadto należy uwzględnić fakt, że jest ona niezależna od wieku.

Istnienie w modelu ścieżek, które mogą przyjmować wartość zerową (tzn. pominięcie pewnych powiązań), wiąże się z problemami dotyczącymi estymacji współczynników ścieżkowych. Możliwe jest bowiem określenie układu dziewięciu jednoczesnych równań, pozwalających na estymację siedmiu wartości współczynników ścieżkowych. Model jest wtedy nadmiernie zidentyfikowany, tzn. uwzględnia zbyt wiele relacji wzajemnie się nakładających i przez to utrudniających właściwą estymację współczynników ścieżkowych. Kendall i O'Muircheartaigh sugerują, aby w takich sytuacjach estymować wszystkie współczynniki w pełni nasyconego modelu i testować istotność tych, które ze względów merytorycznych chcemy usunąć. W rzeczywistości może się okazać, iż współczynniki te są istotne w sensie statystycznym i nie powinny być wyłączone z modelu. Ponieważ zazwyczaj mamy do czynienia ze zbiorowością próbną, postępowanie takie jest uzasadnione.

R. J. A. Little [1980] stosuje analizę ścieżkową do badania wpływu wykształcenia na poziom płodności. W innym modelu rozważa także efekt stosowania środków antykoncepcyjnych. Uwzględniając następujące zmienne: *X1* – wiek, *X2* – wiek w chwili zawarcia małżeństwa, *X3* – stosowanie środków antykoncepcyjnych, *X4* – liczba dzieci, stwierdza, że można je uporządkować przyczynowo w dwojaki sposób:

(I) $$X1 \rightarrow X2 \rightarrow X3 \rightarrow X4$$

(II) $$X1 \rightarrow X2 \rightarrow X4 \rightarrow X3$$

Uporządkowanie przyczynowe pomiędzy antykoncepcją a płodnością w przykładzie (I) wydaje się realne. Można bowiem przypuszczać, iż stosowanie środków antykoncepcyjnych wpływa na liczbę rodzonych przez kobietę dzieci. Jednakże w praktyce częściej spotykanym uporządkowaniem może okazać się wariant (II), szczególnie w krajach o ubogiej tradycji w zakresie planowania rodziny. Oznacza to, że częstość stosowania praktyk antykoncepcyjnych wzrasta wśród kobiet charakteryzujących się większą dzietnością. W konsekwencji płodność okazuje się więc główną determinantą praktyk zapobiegania ciąży, co oczywiście wpływa na dzietność całkowitą.

Omówiony model podkreśla konieczność merytorycznej analizy badanych zagadnień oraz wskazuje na częstą trudność w uporządkowaniu przyczynowym zmiennych. Przyjmując określone wartości zmiennych *X1* i *X2* można oszacować, zgodnie z wersją (I), całkowity wpływ stosowania antykoncepcji na poziom płodności. W wielu populacjach jednak relacja pomiędzy płodnością i antykoncepcją jest „krążąca" – jak w wersji (II). To natomiast pozostaje w sprzeczności z zasadą analizy ścieżkowej mówiącą, że zmienna objaśniana musi być ostatnia w łańcuchu uporządkowania przyczynowego.

Wielu przykładów możliwości zastosowań analizy ścieżkowej w badaniu płodności dostarcza także opublikowana niedawno praca B. Skrętowicz [1991]. Autorka badając uwarunkowania zachowań prokreacyjnych kobiet wiejskich w Polsce, konstruuje modele dla wybranych miar kalendarza rozrodczości, kontroli urodzeń oraz wybranych kategorii dzietności.

Wydaje się zresztą, iż możliwości zastosowań analizy ścieżkowej w demografii są – podobnie jak w socjologii czy psychologii – szczególnie szerokie. Wszędzie tam, gdzie poszukujemy zależności przyczynowo-skutkowych, metoda analizy ścieżkowej może być bardzo użyteczna i pomocna w rozwikłaniu złożonych relacji. Ciesząca się ostatnio bardzo dużą popularnością demografia rodziny – dzięki „naturalnej" sekwencji zdarzeń porządkującej zmienne w sposób przyczynowy – stanowi ogromny obszar badań. Można tu przytoczyć badanie determinant formowania się rodziny, tempa zmian zachodzących w tym procesie, relacji i wzajemności powiązań pomiędzy historią życia rodziny a historią życia jednostek tworzących rodzinę.

Celem niniejszego opracowania, wchodzącego w zakres badań dotyczących płodności kobiet i kształtowania się rodziny, jest próba określenia czynników determinujących te zjawiska. Stawiam hipotezę, iż zastosowanie analizy ścieżkowej pozwoli określonym czynnikom nadać interpretację przyczynową. Poszukiwanie metod badawczych umożliwiających analizę przyczynowości skłania do porównania rezultatów analizy korelacji i regresji z wynikami analizy ścieżkowej. Postępowanie takie powinno udzielić odpowiedzi na pytanie, czy rzeczywiście analiza ścieżkowa stwarza dodatkowe możliwości interpretacji.

2. Źródła statystyczne i charakterystyka badanej zbiorowości

Aplikacja metody wiąże się z poszukiwaniem determinant płodności na podstawie wyników Badania Dzietności Kobiet przeprowadzonego w ramach Narodowego Spisu Powszechnego '88[1]. Omówienie wyników badania dzietności wraz z oceną jakości zebranych danych statystycznych zawiera publikacja J. Paradysza [1992]. Autor, obok analizy struktur demograficznych, powstawania i rozpadu małżeństw, przedstawia również analizę płodności według generacji oraz kohort małżeńskich, a także analizę bieżących odstępstw intergenetycznych.

Przedstawione poniżej wyniki analizy ścieżkowej dotyczą kobiet zamieszkujących województwo poznańskie. Jednostki statystyczne, podobnie jak w całym Badaniu Dzietności, stanowiły kobiety kiedykolwiek zamężne, urodzone w latach 1919 - 1973, zamieszkujące w momencie badania na terenie województwa poznańskiego. Do badania wylosowano 3747 kobiet spełniających powyższe warunki, z czego przebadanych zostało 3098. Odsetek braku odpowiedzi wynosił 17,3%. Był więc zbliżony do wynoszącego 18% braku odpowiedzi w całym badaniu. Znaczny odsetek nieuczestnictwa był zaskoczeniem, tym bardziej iż przedspisowe przewidywania oscylowały w granicach 10%.

Ze względów merytorycznych analizę ograniczono jedynie do kobiet, które zakończyły bądź prawie zakończyły swój okres rozrodczy. Przyjęto umownie, iż należą do nich kobiety urodzone w latach 1919 - 1945, których w przebadanej próbie było 1306, a najmłodsze z nich miały w momencie badania 43 lata. Nie oznacza to oczywiście, iż kobiety w wieku powyżej 43 lat nie rodzą już dzieci, niemniej jednak urodzenia takie należą do rzadkości i nie powinny w sposób zasadniczy wpłynąć na otrzymane rezultaty.

W analizie uwzględniono jedynie cechy demograficzne[2] charakteryzujące natężenie i kalendarz płodności, takie jak: rok urodzenia (wiek kobiety), wiek w chwili ślubu, wiek w chwili rodzenia pierwszego dziecka, liczba urodzonych dzieci oraz okres formowania się rodziny, za który przyjęto odstęp między urodzeniem ostatniego dziecka a datą zawarcia małżeństwa. Chociaż dalszą część analizy ograniczono jedynie do kobiet o zakończonym okresie rozrodczym, warto porównać, jakim zmianom uległy poziom i zróżnicowanie tych cech w dwóch grupach generacji kobiet urodzonych przed i po 1945 roku.

---

[1] Opracowanie niniejsze powstało w ramach badania czynników kształtujących rozwój rodziny i gospodarstwa domowego prowadzonego w ramach grantu KBN nr PB0472/P1/92 „Rozwój rodziny i gospodarstwa domowego".

[2] Szczegółowa analiza płodności kobiet województwa poznańskiego według generacji oraz wszystkich, dostępnych z Badania Dzietności, danych będzie przedmiotem innego, bardziej wyczerpującego opracowania.

Tabela 1 przedstawia wartości średnie i odchylenia standardowe wyróżnionych cech dla obu grup generacji kobiet. Dla określenia poszczególnych cech przyjęto następujące oznaczenia: rok urodzenia (*X1*), wiek w chwili ślubu (*X2*), wiek w chwili rodzenia pierwszego dziecka (*X3*), liczba dzieci (*X4*), długość okresu formowania rodziny (*X5*). Zauważmy, iż nie tylko uległ obniżeniu poziom wyróżnionych cech, lecz także zmniejszyła się ich dyspersja.

Tabela 1

Charakterystyki statystyczne wybranych cech demograficznych kobiet województwa poznańskiego na podstawie badania dzietności kobiet w ramach NSP 1988

Cecha	Generacje 1919 - 1945		Generacje 1946 - 1973			
	$\bar{X}$	S_x	$\bar{X}$	S_x	$\bar{X}'$	S'_x
X1	32,21	7,57	55,33	6,06	51,83	3,82
X2	23,16	4,90	21,68	3,25	21,16	2,68
X3	24,64	4,71	23,00	3,37	22,70	3,08
X4	2,56	1,63	1,99	1,12	2,31	1,16
X5	8,39	4,92	5,95	3,88	6,93	4,02

Uwaga: Symbole zmiennych są objaśnione w tekście, $\bar{X}$ – średnia arytmetyczna, S_x – odchylenie standardowe.

Ze względu na niezakończony okres rozrodczy i ograniczoną porównywalność charakterystyk wyznaczonych dla generacji kobiet 1946 - 1973, wydzielono z nich te kobiety, które zawarły związek małżeński po 1980 r. Wyznaczone dla takiej grupy charakterystyki oznaczono symbolami: $\bar{X}'$, S'_x.

Źródło: Obliczenia własne na podstawie danych Katedry Statystyki i Demografii Akademii Ekonomicznej w Poznaniu z Badania Dzietności Kobiet w ramach NSP 1988.

Przede wszystkim zmniejszyła się dzietność kobiet kiedykolwiek zamężnych (*X4*) z 2,56 do 1,99 dziecka. Niezakończony okres rozrodczy i krótki staż małżeński, kobiet z grupy generacji 1946 - 1973, w dużej mierze wpłynęły na deformacje charakterystyk natężenia i kalendarza płodności. Dlatego dla celów porównawczych wyeliminowano z tej grupy kobiety, które wyszły zamąż po 1980 roku. Wówczas spadek dzietności jest zdecydowanie mniejszy, bo do 2,31. Wyraźne jest również zmniejszenie dyspersji z 1,63 do 1,12 (bądź do 1,16 – jeśli pominąć kobiety, które zawarły związek małżeński po 1980 roku). Sytuacja ta dotyczy nie tylko dzietności, ale również pozostałych cech badanych kobiet. Przeciętny wiek zamążpójścia (*X2*) obniżył się z 23,16 do 21,68 (21,16) lat. Wiek w chwili rodzenia pierwszego dziecka (*X3*) zmniejszył się w młodszych generacjach kobiet z 24,64 do 23,0 (22,7). Jednocześnie zmniejszyła się dyspersja tego wieku z 4,71 do 3,37 (3,08). Okres formowania rodziny (*X5*) również uległ skróceniu, choć kwestia porównań tego ostatniego może budzić pewne wątpliwości. Zaobserwowane różnice (por. tab. 1) są we wszystkich przypadkach wyraźne.

3. Analiza współzależności

W analizie regresji model przedstawiający zależność liniową określony jest zazwyczaj następującym równaniem:

$$X_0 = a_{01}X_1 + a_{02}X_2 + \cdots + a_{0k}X_k + a_{0w}X_w. \tag{1}$$

Reprezentuje on bezpośredni wpływ wybranych k zmiennych objaśniających na zmienną objaśnianą. Interpretując wartości cząstkowych współczynników regresji zakłada się, iż pozostałe zmienne nie ulegają zmianom. Podstawową zaletą analizy ścieżkowej jest to, że dzięki dekompozycji powyższego równania umożliwia ona dalsze rozpoznanie „mechanizmu zależnościowego". Zakłada się, że zmiana wartości danej zmiennej – nie powodując zmian zmiennych wcześniejszych w uporządkowaniu przyczynowym – może poprzez zmiany zmiennych po niej następujących wpływać na wartość zmiennej objaśnianej. Opierając się na diagramie ścieżkowym[3] skonstruowanym na podstawie merytorycznej analizy badanej zależności, równanie (1) uzupełniamy o kolejne, przedstawiające wszystkie możliwe (ujęte) na diagramie relacje:

$$\begin{aligned} X_2 &= a_{21}X_1 + a_{2u}X_u, \\ X_3 &= a_{31}X_1 + a_{32}X_2 + a_{3v}X_v, \\ &\cdots\cdots\cdots\cdots\cdots \\ X_k &= a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{k(k-1)}X_{k-1} + a_{kz}X_z. \end{aligned} \tag{2}$$

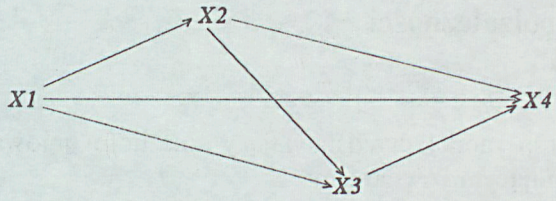
Rozważmy wpływ wieku zamążpójścia kobiet na ich płodność (model I). Powszechnie przyjmuje się, iż im niższy wiek kobiet w chwili ślubu, tym niższy wiek w momencie rodzenia pierwszego dziecka, tym dłuższy okres narażenia na ryzyko poczęcia, tym większa dzietność. Spróbujmy uporządkować przyczynowo wyżej wymienione czynniki według ich wzajemnych relacji oraz sposobu oddziaływania na płodność.

WIEK → WIEK W CHWILI ZAWARCIA MAŁŻEŃSTWA → WIEK W CHWILI RODZENIA PIERWSZEGO DZIECKA → LICZBA DZIECI

Odpowiedni diagram ścieżkowy przyjmie w tym przypadku następującą postać (rys. 1):

---

[3] Zasady konstrukcji diagramu ścieżkowego oraz metody szacunku współczynników ścieżkowych zostały omówione w pracy E. Gołata [1992].



Rys. 1. Diagram ścieżkowy w analizie wpływu czterech zmiennych

Zazwyczaj pierwszym etapem w analizie współzależności jest analiza korelacji. Przeprowadza się ją na podstawie macierzy współczynników korelacji liniowej, K, która w rozważanym przykładzie jest następująca:

$$K=\begin{bmatrix} 1 & -0{,}227 & -0{,}204 & -0{,}233 \\ -0{,}227 & 1 & 0{,}743 & -0{,}142 \\ -0{,}204 & 0{,}743 & 1 & -0{,}233 \\ -0{,}233 & -0{,}142 & -0{,}233 & 1 \end{bmatrix}$$

Wartości wszystkich współczynników korelacji są istotne w sensie statystycznym dla $\alpha=0{,}001$. Jednakże, jak wyraźnie widać, nie są wysokie co do wartości bezwzględnej (z wyjątkiem r_{23}), co świadczy o raczej słabej zależności korelacyjnej pomiędzy wyróżnionymi zmiennymi.

Interesującą nas relację pomiędzy wiekiem zamążpójścia a dzietnością można przedstawić za pomocą współczynnika korelacji pomiędzy zmiennymi określającymi obydwa czynniki $r_{24}=-0{,}142$. Oznacza to, iż zróżnicowanie dzietności zaledwie w 2% zdeterminowane jest przez wiek kobiet w chwili ślubu, przy czym wzrost wieku zamążpójścia wiąże się ze spadkiem dzietności. Jednakże chcąc określić wpływ wieku w chwili ślubu na płodność przy uwzględnieniu wieku w momencie rodzenia pierwszego dziecka, musimy się posłużyć współczynnikiem korelacji cząstkowej pierwszego rzędu, $r_{24.3}=0{,}057$, bądź wyższego, jeśli interesuje nas dodatkowo uwzględnienie innych zmiennych np. $r_{24.13}=0{,}025$.

W dalszej kolejności przeprowadza się zazwyczaj analizę regresji. Zależność liniową pomiędzy wyróżnionymi cechami można przedstawić w postaci następującego równania regresji wielu zmiennych[4]:

[4] Modele zostały przedstawione w formie zgodnej z ogólnie przyjętymi zasadami zapisu modeli regresyjnych. W nawiasach podano wartości błędów standardowych szacunku odpowiednich parametrów strukturalnych. Poniżej znajdują się wartości statystyki t-Studenta umożliwiające ocenę istotności wpływu poszczególnych zmiennych. Dla oceny jakości modeli podano także wartości współczynnika determinacji R^2 oraz odchylenia standardowego składnika resztowego S_z.

$$X4 = \underset{(0,005)}{-0,055}\, X1 + \underset{(0,012)}{0,012}\, X2 - \underset{(0,012)}{0,122}\, X3 + Z4, \qquad (3)$$
$$\begin{array}{ccc} -10,185 & 0,953 & -10,347 \end{array}$$

$R^2 = 0,167,$
$S_{Z4} = 1,400.$

Uwzględnione w równaniu regresji zmienne objaśniające zaledwie w 17% wyjaśniają zmienność dzietności. Otrzymane rezultaty informują, że im starsza generacja, tym przeciętnie o 0,05 większa dzietność kobiet, przy założeniu stałości pozostałych warunków. Podobnie wzrost wieku kobiet o rok, w momencie rodzenia pierwszego dziecka ($X3$), wiąże się średnio z obniżeniem dzietności o 0,12.

Kierunek oddziaływania wieku kobiet w chwili ślubu ($X2$) jest natomiast dyskusyjny. Przeciwnie, w stosunku do warunków analizy korelacji, cząstkowy współczynnik regresji charakteryzuje się ujemnym znakiem. Warto zwrócić uwagę na fakt, iż właśnie ta ostatnia zmienna okazuje się nieistotna w sensie statystyki t-Studenta. Współczynnik regresji $b_{42.13}$ charakteryzuje się dużym błędem standardowym szacunku i zmienną $X2$ należałoby usunąć z modelu.

Współczynniki regresji a_{ij} liniowego równania regresji, bądź układu równań (2), można interpretować jako miary zależności przyczynowych. Są to miary zależne od jednostek, w jakich dokonano pomiaru zmiennych X_i i X_j. Współczynniki a_{ij} mogą przyjmować dowolne wartości rzeczywiste. Analizę ścieżkową można przeprowadzić na podstawie zarówno rzeczywistych, jak i unormowanych wartości współczynników regresji. Brak normalizacji, między innymi, odróżnia współczynniki regresji od współczynników korelacji. Oczywiście to, iż współczynnik jest miarą „rzeczywistą" czy „abstrakcyjną" nie przesądza o jego wartości. Współczynnik regresji wyraża zmiany zmiennej zależnej spowodowane jednostkowym przyrostem zmiennej niezależnej. Pewne sytuacje badawcze mogą wymagać takiej właśnie miary zależności między zmiennymi. Miara ta, jednak nie w każdym przypadku, nadaje się do dokonywania porównań.

Jeżeli zmienne X zastąpimy przez odpowiadające im zmienne standaryzowane ($U_i = (X_i - \bar{X}_i)/S_i$), to analizowane równanie (3) przyjmie następującą postać:

$$U4 = \underset{(0,025)}{-0,257}\, U1 + \underset{(0,034)}{0,035}\, U2 - \underset{(0,034)}{0,354}\, U3 + E4, \qquad (4)$$
$$\begin{array}{ccc} -10,185 & 0,953 & -10,347 \end{array}$$

$R^2 = 0,167,$
$S_{E4} = 0,860.$

Postępując zgodnie z ideą analizy ścieżkowej, równanie (4) uzupełniamy o następujący układ równań obrazujący przedstawione na diagramie ścieżkowym (rys. 1) relacje:

$$U3 = \underset{(0{,}021)}{-0{,}031}\, U1 + \underset{(0{,}022)}{0{,}737}\, U2 + E3.$$
$$\begin{matrix} & (0{,}021) & (0{,}022) \\ & -1{,}456 & 32{,}705 \end{matrix}$$

$R^2 = 0{,}483,$
$S_{E3} = 0{,}720,$

(5)

$$U2 = \underset{(0{,}027)}{-0{,}188}\, U1 + E2,$$
$$\begin{matrix} & (0{,}027) \\ & -6{,}910 \end{matrix}$$

$R^2 = 0{,}035,$
$S_{E2} = 0{,}980.$

Współczynniki regresji p_{ij} otrzymane dla zmiennych standaryzowanych noszą nazwę współczynników ścieżkowych, a rozwiązanie i interpretacja odpowiedniego układu równań nazywana jest analizą ścieżkową.

Współczynnik ścieżkowy jest miarą zależności zmian U_i od zmian U_j. Informuje on o tym, jaka część odchylenia standardowego zmiennej zależnej U_i występuje wtedy, gdy zmienna niezależna U_j ulega zmianom, a pozostałe czynniki pozostają stałe. Współczynnik p_{ij} przybiera wartość 1, gdy cała zmienność U_i może być przewidziana na podstawie zmienności zmiennej U_j. Analiza ścieżkowa odróżnia korelację, jako relację symetryczną, od stopnia, w jakim jedna ze zmiennych wpływa na drugą, zakładając asymetrię tego wpływu.

Współczynniki ścieżkowe przypominają współczynniki korelacji w opisie relacji między zmiennymi w „abstrakcyjnej" skali. Diagram przedstawiający relacje między zmiennymi i jednocześnie przyporządkowujący każdemu wektorowi wartość odpowiedniego współczynnika ścieżkowego udziela natychmiast informacji o względnym wpływie zmienności czynników bezpośrednich w uporządkowaniu przyczynowym na zmienność poszczególnych zmiennych.

Współczynniki ścieżkowe różnią się od współczynników korelacji tym, że mogą przybierać wartości spoza przedziału $\langle -1, 1\rangle$. Wartość współczynnika ścieżkowego większa od jedności wskazuje natychmiast, że bezpośrednie oddziaływanie czynnika powoduje większą zmienność (dyspersję zmiennej) aniżeli aktualnie obserwowana. Bezpośredni efekt jest wtedy konsekwencją nakładających się wpływów innych czynników.

Dla systemu, w którym nie występują relacje zwrotne, przy założeniu poprawnie skonstruowanego diagramu ścieżkowego, można sformułować następującą zasadę wyznaczania korelacji[5] pomiędzy dwiema jakimikolwiek

[5] Współczynniki ścieżkowe można wyznaczyć również metodą dekompozycji współczynnika korelacji według następującej formuły: $r_{ij} = \sum_q p_{iq} r_{qi}$, gdzie q zmienia się według wszystkich zmiennych, przez które przechodzi ścieżka. Każdy ze składników powyższego równania ma interpretację merytoryczną jako prosty bądź złożony współczynnik ścieżkowy, odzwierciedlający bezpośredni bądź pośredni wpływ zmiennych przyczynowo wcześniejszych na zmienną objaśnianą.

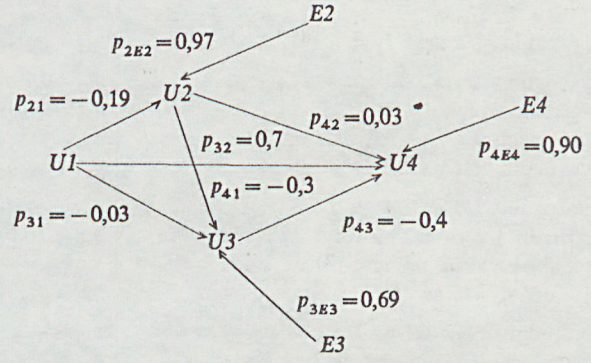
zmiennymi. Głosi ona, że współczynnik korelacji jest równy sumie wpływów dochodzących do ścieżki, według której można śledzić oddziaływanie jednej zmiennej na drugą.

Po kroku w przód nie należy się cofać i dwa razy przechodzić przez tę samą zmienną w jednej ścieżce. Współczynnik odpowiadający całej ścieżce łączącej dwie zmienne i mierzący udział całej ścieżki w danej korelacji nazywany jest ścieżkowym współczynnikiem złożonym. Jego wartość wyznacza się jako iloczyn współczynników odpowiadających ścieżkom elementarnym wyznaczającym ścieżkę główną. Tylko jedna z tych ścieżek może być ścieżką dwustrzałkową (odpowiadającą współczynnikowi korelacji), bez pogwałcenia zasady cofania się po uprzednim kroku w przód.

Spełnienie warunku całkowitej determinacji wymaga uzupełnienia modelu o resztowe współczynniki ścieżkowe[6]. Przyjmują one następujące wartości:

$$p_{2E2}=0{,}97, \quad p_{3E3}=0{,}69, \quad p_{4E4}=0{,}90.$$

Ostatecznie więc kompletny diagram ścieżkowy można przedstawić następująco:



Rys. 2. Diagram ścieżkowy w analizie oddziaływania wieku kobiet, wieku w chwili ślubu i wieku w momencie rodzenia pierwszego dziecka na dzietność kobiet

4. Omówienie wyników

Zgodnie z wcześniejszymi rozważaniami można stwierdzić, że globalny efekt wpływu wieku zamążpójścia (*U2*) na płodność (*U5*) jest uzależniony od wieku respondentki (*U1*), ale nie od wieku w chwili rodzenia pierwszego dziecka (*U3*).

[6] Wartości ścieżkowych współczynników resztowych można wyznaczyć na podstawie formuły całkowitej determinacji: $r_{ii}=1=\sum p_{iq}r_{qi}$, która po odpowiednim rozpisaniu i przekształceniu daje wzór: $p^2_{iEi}=1-\sum p_{iq}r_{qi}$. Por. P. Klissou [1991].

Oznacza to, iż zmiana wieku zamążpójścia (*U2*) nie powoduje zmiany w strukturze wieku populacji (*U1*), ale oddziałuje na płodność (*U4*), przede wszystkim skracając okres trwania małżeństwa (ryzyka poczęcia), tj. poprzez opóźnienie wieku rodzenia dzieci (*U3*). Jest to zgodne z zasadą analizy ścieżkowej, która głosi, iż zmiana wartości danej zmiennej nie wpływa na wartość zmiennej wcześniejszej w uporządkowaniu przyczynowym. Jedynie poprzez wpływ na wartość następnej zmiennej w łańcuchu uporządkowania przyczynowego badana zmienna może oddziaływać na wartość zmiennej objaśnianej.

Omówienie rezultatów przeprowadzonej analizy ścieżkowej ułatwi przedstawienie ich w postaci tabeli 2.

Tabela 2

Ścieżkowa dekompozycja wpływów w analizie oddziaływania roku urodzenia kobiet (*U1*), wieku w chwili ślubu (*U2*) i wieku w momencie rodzenia pierwszego dziecka (*U3*) na dzietność kobiet (*U4*)

Zmienna objaśniana	Charakter oddziaływania	Współczynnik ścieżkowy
U2	wpływ ogólny *U1* na *U2*	$p_{21} = -0{,}2$
U3	wpływ ogólny *U2* na *U3*	$p_{32} = 0{,}7$
	bezpośredni wpływ *U1* na *U3*	$p_{31} = -0{,}03$
	pośredni wpływ *U1* poprzez *U2* na *U3*	$p_{32}p_{21} = -0{,}14$
	wpływ ogólny *U1* na U3	$p_{31} + p_{32}p_{21} = -0{,}2$
U4	wpływ ogólny *U3* na *U4*	$p_{43} = -0{,}4$
	bezpośredni wpływ *U2* na *U4*	$p_{42} = 0{,}03$
	pośredni wpływ *U2* poprzez *U3* na *U4*	$p_{43}p_{32} = -0{,}21$
	wpływ ogólny *U2* na *U4*	$p_{42} + p_{43}p_{32} = -0{,}19$
	bezpośredni wpływ *U1* na *U4*	$p_{41} = -0{,}3$
	pośredni wpływ *U1* poprzez *U2* na *U4*	$p_{42}p_{21} = -0{,}006$
	pośredni wpływ *U1* poprzez *U3* na *U4*	$p_{43}p_{31} = 0{,}012$
	pośredni wpływ *U1* poprzez *U2* i *U3* na *U4*	$p_{43}p_{32}p_{21} = 0{,}056$
	wpływ ogólny *U1* na *U4*	$p_{41} + p_{43}p_{31} + p_{42}p_{21} + p_{43}p_{32}p_{21} = -0{,}23$

Źródło: Obliczenia własne na podstawie danych Katedry Statystyki i Demografii Akademii Ekonomicznej w Poznaniu z Badania Dzietności Kobiet w ramach NSP 1988.

Okazuje się przede wszystkim – tak jak sugerowały to już wyniki analizy korelacji i regresji – iż ogólne oddziaływanie wieku zamążpójścia na dzietność kobiet jest niezbyt silne ($p_{42} + p_{43}p_{32} = -0{,}19$). Oznacza to, że zmienność wieku w chwili ślubu w około 19% wyjaśnia zróżnicowanie dzietności. Znak złożonego współczynnika ścieżkowego jest ujemny, a więc zgodny ze znakiem współczynnika korelacji prostej, jednakże odmienny aniżeli znak cząstkowego współ-

czynnika regresji. Ta zmienność kierunku oddziaływania wiąże się z dezagregacją wpływu ogólnego zmiennej *U2* na *U4*. Bezpośredni wpływ wieku zamążpójścia na dzietność kobiet jest dodatni i bliski zeru $p_{42}=0{,}03$. Natomiast pośrednie oddziaływanie *U2* na *U4* poprzez podwyższanie wieku w momencie rodzenia pierwszego dziecka *U3* i skracanie w ten sposób efektywnego okresu rozrodczego ujemnie wpływa na dzietność *U4*. Odpowiedni, złożony współczynnik ścieżkowy

$$p_{432}=p_{43}p_{32}=-0{,}3\cdot 0{,}7=-0{,}21$$

wskazuje, iż wpływ pośredni jest zdecydowanie silniejszy aniżeli bezpośredni. Tak więc wpływ ogólny *U2* na *U4* jest wypadkową wpływów bezpośredniego i pośredniego.

Ujemny ogólny wpływ wieku kobiet (w sensie roku urodzenia), $r_{14}=-0{,}23$, oznacza, iż starsze generacje kobiet charakteryzują się wyższą dzietnością. Ciekawa jest jego dekompozycja wskazująca na bezpośrednie i pośrednie oddziaływanie poprzez uwzględnione w badaniu zmienne. Bezpośredni wpływ wieku $p_{41}=-0{,}3$ jest ujemny i wyższy co do wartości bezwzględnej od odpowiedniego współczynnika korelacji. Informuje on, iż w 30% zmienność dzietności jest wyjaśniona bezpośrednio przez rok urodzenia kobiety. W tym samym kierunku działa pośrednie oddziaływanie *U1* poprzez *U2* na *U4*. Złożony współczynnik ścieżkowy

$$p_{421}=p_{42}p_{21}=-0{,}006$$

przyjmuje jednak wartość bardzo bliską zeru. Śledząc natomiast pośrednie oddziaływanie *U1* poprzez *U3*

$$p_{431}=p_{43}p_{31}=0{,}012$$

oraz poprzez *U2* i *U3*

$$p_{4321}=p_{43}p_{32}p_{21}=0{,}056,$$

otrzymujemy informację o dadatnim, pozytywnym kierunku wpływu powodującym niewielki wzrost dzietności. Można to tłumaczyć następująco: im starsza generacja, tym wyższy wiek zamążpójścia, tym wyższy wiek w chwili rodzenia pierwszego dziecka, tym mniejsza dzietność. Ten „wzrost"płodności koryguje jej spadek związany z przechodzeniem do coraz to młodszych generacji i ostatecznie ogólny wpływ wieku na dzietność wynosi

$$p_{41}+p_{43}p_{31}+p_{42}p_{21}+p_{43}p_{32}p_{21}=-0{,}23.$$

Warto zauważyć, iż ogólny wpływ zmiennej *U3* na zmienną *U4* nie jest równy współczynnikowi korelacji prostej $r_{43}=-0{,}23$. Ogólny wpływ wieku w chwili rodzenia pierwszego dziecka na dzietność to wpływ bezpośredni $p_{43}=-0{,}4$, ponieważ w uporządkowaniu przyczynowym nie uwzględniono żadnych zmiennych pomiędzy *U3* a *U4*. Wartości współczynników korelacji

r_{43} oraz ścieżkowego p_{43} sugerują, że część wpływu przypisanego zmiennej *U3* spowodowana została przez oddziaływanie zmiennych przyczynowo wcześniejszych, przede wszystkim wieku w chwili ślubu.

Jak zauważono wcześniej, zmienna *U2* jest nieistotna w sensie statystyki *t*-Studenta i dlatego należałoby ją usunąć z modelu. Usunięcie zmiennej *U2* powoduje oczywiście poprawę modelu. Wartości parametrów strukturalnych przy pozostałych zmiennych nie ulegają zmianie, co oznacza ich stacjonarność. Wartość współczynnika korelacji wielorakiej ulega obniżeniu zaledwie o 0,01. Tak więc pod względem formalnym postępowanie takie jest uzasadnione. Jednakże warto pamiętać, że usunięcie zmiennej z modelu redukuje liczbę informacji możliwych do otrzymania. W tym przypadku pozbawia nas informacji o sposobie oddziaływania wieku w chwili ślubu na dzietność kobiet.

Interesujące wydaje się rozważenie jeszcze jednej zależności (model II), tym razem z pominięciem zmiennej *U2*, a mianowicie wpływu wyróżnionych zmiennych na długość okresu formowania rodziny (*U5*). Uporządkowanie przyczynowe zmiennych można określić w następujący sposób:

WIEK → WIEK W CHWILI RODZENIA PIERWSZEGO DZIECKA →
→ LICZBA DZIECI → DŁUGOŚĆ OKRESU FORMOWANIA RODZINY

Macierz współczynników korelacji prostej pomiędzy zmiennymi, L, przedstawia się następująco:

$$\mathrm{L}=\begin{bmatrix} 1 & -0{,}204 & -0{,}233 & -0{,}072 \\ -0{,}204 & 1 & -0{,}233 & -0{,}120 \\ -0{,}233 & -0{,}233 & 1 & 0{,}551 \\ -0{,}072 & -0{,}120 & 0{,}551 & 1 \end{bmatrix}.$$

Wszystkie wartości współczynników korelacji liniowej, z wyjątkiem r_{15}, są istotne w sensie statystycznym. Najsilniejsza jest jednak zależność korelacyjna długości okresu formowania rodziny od dzietności $r_{45}=0{,}551$. Oznacza to, że zwiększenie liczby dzieci (*U4*) powoduje wydłużenie okresu kształtowania rodziny (*U5*). Natomiast współczynnik korelacji pomiędzy wiekiem w chwili rodzenia pierwszego dziecka a zmienną *U5* jest niższy co do wartości bezwzględnej i ujemny. Tak więc im niższy wiek rodzenia pierwszego dziecka, tym dłuższy okres tworzenia rodziny.

Zbudowany dla powyższych zależności model ścieżkowy przyjmuje następującą postać:

$$\begin{array}{lllll} U5=0{,}069\,U1 & +0{,}037\,U3 & +0{,}650\,U4 & +E5, & (6) \\ \quad (0{,}028) & \quad (0{,}032) & \quad (0{,}032) & & \\ \quad 2{,}46 & \quad 1{,}15 & \quad 20{,}31 & & \end{array}$$

$R^2=0{,}308,$
$S_{E5}=0{,}833,$

$$U4 = -0{,}260\,U1 - 0{,}332\,U3 + E4,$$
$$\begin{matrix} (0{,}025) & (0{,}025) \\ -10{,}376 & -13{,}285 \end{matrix}$$

$R^2 = 0{,}167,$
$S_{E4} = 0{,}863,$

(7)

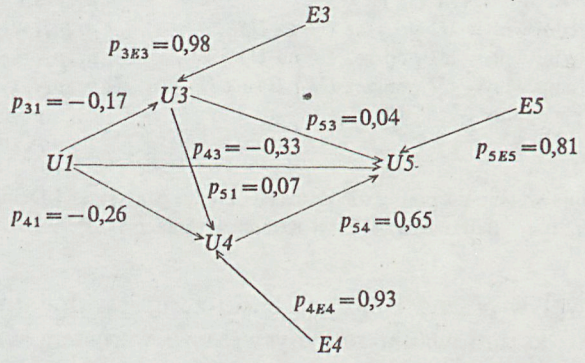
$$U3 = -0{,}172\,U1 + E3,$$
$$\begin{matrix} (0{,}028) \\ -6{,}076 \end{matrix}$$

$R^2 = 0{,}029,$
$S_{E3} = 0{,}986.$

Resztowe współczynniki ścieżkowe dla powyższego modelu równe są odpowiednio:

$$p_{3E3} = 0{,}98, \quad p_{4E4} = 0{,}93, \quad p_{5E5} = 0{,}81.$$

Natomiast kompletny diagram ścieżkowy przedstawiono na rysunku 3.



Rys. 3. Diagram ścieżkowy w analizie oddziaływania wieku kobiet, wieku w momencie rodzenia pierwszego dziecka oraz liczby dzieci na długość okresu formowania rodziny

Wartości zarówno prostych, jak i złożonych współczynników ścieżkowych, które przedstawiono w tabeli 3, umożliwiają wyciągnięcie poniższych wniosków. Przede wszystkim zauważmy, że wartości prostych współczynników ścieżkowych, odzwierciedlających wpływy bezpośrednie na zmienną *U5* (współczynników regresji w równaniu (6) dla zmiennych standaryzowanych), są dodatnie. Natomiast współczynniki korelacji r_{15} i r_{13} przyjmują wartości ujemne. Spróbujmy wyjaśnić te rozbieżności.

Ogólne oddziaływanie wieku w chwili rodzenia pierwszego dziecka na długość okresu formowania rodziny jest – zgodnie z wynikami analizy korelacji – ujemne

$$p_{53} + p_{54}p_{43} = -0{,}18.$$

Tabela 3

Ścieżkowa dekompozycja wpływów w analizie oddziaływania roku urodzenia kobiet (*U1*), wieku w momencie rodzenia pierwszego dziecka (*U3*) oraz liczby dzieci (*U4*) na długość okresu formowania rodziny (*U5*)

Zmienna objaśniana	Charakter oddziaływania	Współczynnik ścieżkowy
U3	wpływ ogólny *U1* na *U3*	$p_{31} = -0{,}17$
U4	wpływ ogólny *U3* na *U4*	$p_{43} = -0{,}33$
	bezpośredni wpływ *U1* na *U4*	$p_{41} = -0{,}26$
	pośredni wpływ *U1* poprzez *U3* na *U4*	$p_{43}p_{31} = 0{,}06$
	wpływ ogólny *U1* na U4	$p_{41} + p_{43}p_{31} = -0{,}2$
U5	wpływ ogólny *U4* na *U5*	$p_{54} = 0{,}65$
	bezpośredni wpływ *U3* na *U5*	$p_{53} = 0{,}04$
	pośredni wpływ *U3* poprzez *U4* na *U5*	$p_{54}p_{43} = -0{,}22$
	wpływ ogólny *U3* na *U5*	$p_{54} + p_{54}p_{43} = -0{,}18$
	bezpośredni wpływ *U1* na *U5*	$p_{51} = 0{,}7$
	pośredni wpływ *U1* poprzez *U3* na *U5*	$p_{53}p_{31} = -0{,}01$
	pośredni wpływ *U1* poprzez *U4* na *U5*	$p_{54}p_{41} = -0{,}17$
	pośredni wpływ *U1* poprzez *U3* i *U4* na *U5*	$p_{54}p_{43}p_{31} = 0{,}037$
	wpływ ogólny *U1* na *U5*	$p_{51} + p_{53}p_{31} + p_{54}p_{41} + p_{54}p_{43}p_{31} = -0{,}07$

Źródło: Obliczenia własne na podstawie danych Katedry Statystyki i Demografii Akademii Ekonomicznej w Poznaniu z Badania Dzietności Kobiet w ramach NSP 1988.

Bezpośredni wpływ *U3* na *U5*, choć bliski zeru, jest dodatni, co oznacza wydłużenie okresu kształtowania rodziny wraz ze wzrostem wieku rodzenia pierwszego dziecka – bądź odwrotnie. Pośrednie oddziaływanie *U3* poprzez *U4* na *U5* jest natomiast ujemne i znacznie silniejsze co do wartości bezwzględnej

$$p_{543} = p_{54}p_{43} = 0{,}65 \cdot (-0{,}332) = -0{,}22.$$

Interpretacja tej zależności może być następująca: im wyższy wiek rodzenia pierwszego dziecka (*U3*), tym mniej dzieci (*U4*), im mniej dzieci, tym krótszy okres kształtowania rodziny. Ponieważ pośrednie oddziaływania *U3* na *U5* poprzez *U4* jest silniejsze aniżeli wpływ bezpośredni, ostateczny efekt (choć umniejszony) wskazuje, że wcześniejszy wiek rozpoczęcia formowania rodziny powoduje wydłużenie tego okresu. Ostatecznie zmienność *U3* w około 18% wyjaśnia zróżnicowanie zmiennej *U5*.

Najlepiej wyjaśnia analiza ścieżkowa wpływ wieku kobiet na długość formowania rodziny. Choć otrzymane wartości współczynników ścieżkowych

są niewysokie, warto je przeanalizować. Im starsza generacja (mniejsza wartość *U1*), tym dłuższy okres kształtowania rodziny (*U5*) $p_{51}=0{,}07$. Bezpośredni wpływ, w przeciwieństwie do współczynnika korelacji, jesdodatni. Oddziaływanie *U1* na *U5* poprzez *U3*

$$p_{531}=p_{53}p_{31}=0{,}037\cdot(-0{,}172)=-0{,}01,$$

choć wyjaśnia zaledwie 1% zmienności *U5*, mówi, że im starsza generacja, tym wyższy wiek rodzenia pierwszego dziecka, tym dłużej kształtuje się rodzina. Z kolei wpływ *U1* na *U5* poprzez *U4*

$$p_{541}=p_{54}p_{41}=0{,}65\cdot(-0{,}26)=-0{,}169$$

informuje, że im młodsza generacja, tym mniej dzieci, tym krócej formuje się rodzina. To pośrednie oddziaływanie wieku poprzez liczbę dzieci wyjaśnia blisko 17% zmienności *U5*. Pośredni wpływ *U1* na *U5* poprzez *U3* i *U4* okazuje się dodatni

$$p_{5431}=p_{54}p_{43}p_{31}=0{,}65\cdot(-0{,}332)\cdot(-0{,}172)=0{,}037.$$

Tak więc im młodsza generacja, tym niższy wiek rodzenia pierwszego dziecka, tym więcej dzieci, tym dłużej kształtuje się rodzina. Ostatecznie jednak ogólny wpływ wieku jest równy bezpośredniemu, co do wartości bezwzględnej, ale o przeciwnym znaku, ponieważ oddziaływania pośrednie są silniejsze

$$p_{51}+p_{53}p_{31}+p_{54}p_{41}+p_{54}p_{43}p_{31}=-0{,}07.$$

Najsilniejsze, jak należało przypuszczać, jest oddziaływanie zmiennej *U4* na *U5*. Odpowiedni współczynnik ścieżkowy przyjmuje wartość $p_{54}=0{,}65$. Wzrost liczby dzieci w ponad 60% wyjaśnia więc zróżnicowanie okresu kształtowania rodziny.

Uwagi końcowe

Wśród demograficznych uwarunkowań dzietności wiek zamążpójścia uznawany jest za jedno z istotniejszych (model I). Tymczasem wyniki analizy korelacji i regresji uzyskane na podstawie danych z Badania Dzietności dla kobiet województwa poznańskiego nie potwierdzają tego stwierdzenia. Informują one o niezbyt wysokiej sile związku, a odmienne znaki odpowiednich współczynników sugerują nawet zjawisko koincydencji.

Analiza ścieżkowa poprzez dezagregację wpływu ogólnego umożliwia szczegółowe przeanalizowanie zawikłanego mechanizmu zależnościowego. Okazuje się, iż bezpośrednie oddziaływanie wieku zamążpójścia na płodność jest faktycz-

nie dodatnie, aczkolwiek minimalne co do siły. Natomiast wpływ pośredni, poprzez podwyższenie wieku w momencie rodzenia pierwszego dziecka i skracanie efektywnego okresu rozrodczego, jest ujemny i zdecydowanie silniejszy.

Również w drugim z rozpatrywanych modeli wpływ wieku w chwili rodzenia pierwszego dziecka na długość okresu formowania rodziny, jak sugeruje analiza korelacji i regresji, jest nieistotny. Dopiero podejście analizy ścieżkowej wyjaśnia to oddziaływanie w kategoriach wpływu pośredniego poprzez inne czynniki.

Szczegółowa analiza uwarunkowań dzietności wymaga oczywiście odrębnego, szerszego opracowania. Należy przede wszystkim uwzględnić podział kobiet według generacji i miejsca zamieszkania. Należy także rozszerzyć zakres rozpatrywanych czynników. Natomiast uwzględnienie analizy ścieżkowej wśród stosowanych metod badania przyczynowości, jak wykazano, może okazać się bardzo pomocne w wyjaśnianiu skomplikowanych, rzeczywistych relacji pomiędzy zmiennymi demograficznymi. Stwarza ona szerokie możliwości interpretacji, umożliwiające przeprowadzenie pogłębionej analizy statystyczno-demograficznej.

Literatura

Blalock, H. M. [1964], Causal Inferences in Nonexperimental Research, Chapel Hill.

Duncan, O. D. [1966], Path Analysis: Sociological Exemples. „The American Journal of Sociology" Vol. 72, No. 1.

Gołata, E. [1992], Możliwości zastosowania analizy ścieżkowej w demografii, „Studia Demograficzne", nr 4.

Karpiński, J. [1985], Przyczynowość w badaniach socjologicznych, PWN, Warszawa.

Kendall, M. G., O'Muircheartaigh, C. A. [1977], Path Analysis and Model Building, WFS, Technical Bulletins, No. 2/Tech. 414.

Klissou, P. [1991], La multicollinéarité dans l'analyse des cheminements. Un essai d'application a'la démographie, Institut de Démographie, Université Catholique de Louvain, Louvain-la-Neuve, Working Paper No. 160.

Little, R. J. A. [1980], Linear Models and Path Analysis, WFS, Occasional Papers, No. 22.

Paradysz, J. [1992], Dzietność kobiet w Polsce, Materiały i opracowania statystyczne NSP'88, GUS, Warszawa.

Skrętowicz, B. [1991], Uwarunkowania zachowań prokreacyjnych kobiet wiejskich. Monografie i Opracowania, Nr 356/28, ISiD, Szkoła Główna Handlowa.

Wright, S. [1934], The Method of Path Coefficients, „Annals of Mathematical Statistics, V.

Wright, S. [1960], Path Coefficients and Path regressions: Alternative or Complementary Concepts?, „Biometrics", No. 16.