Wiktor Ejsmont
Janusz Łyko
Uniwersytet Ekonomiczny we Wrocławiu

WPŁYW POŁOŻENIA SZKOŁY NA WYNIKI EDUKACYJNE UCZNIÓW*

Wprowadzenie

Główne zmiany organizacyjne w polskim szkolnictwie po 1989 r. sprowadzają się do wprowadzenia trzystopniowego nauczania oraz odejścia na etapie szkół ponadgimnazjalnych od szkolnictwa zawodowego na rzecz nauczania w liceach ogólnokształcących. Zmianie uległy także wzorce i stereotypy dotyczące miejsca zamieszkania. Rosnące bezrobocie oraz rozwój rynku nieruchomości spowodowały coraz częstsze migracje ludności. Migracje te wpłynęły także na proces edukacji młodego pokolenia. Przemieszczając się wraz z rodzicami dzieci zaczęły coraz częściej zmieniać swoje środowisko szkolne. W związku z tym istotną rolę zaczął odgrywać problem lokalizacji różnego poziomu szkół. Przy tej okazji rodzi się pytanie o ocenę poszczególnych placówek w połączeniu z ich lokalizacją.

Główna cześć pracy opiera się na zastosowaniu modelu efektywności nauczania, który został opisany przez M. Aitkina i N. Longforda w artykule „Statistical Modelling Issues in School Effectiveness Studies". Przedstawiono tam kilka modeli służących do badania efektywności kształcenia w amerykańskich szkołach za pomocą różnicy punktów pomiędzy wynikami egzaminów osiąganymi przez uczniów kończących szkołę a ilorazem inteligencji IQ, mierzonym przed rozpoczęciem nauki w danej szkole. Celem prezentowanego artykułu jest analiza przyrostu wiedzy uczniów w zależności od miejsca ukończenia gimnazjum oraz liceum. W tym celu wykorzystano model z losowymi efektami − rozpatrywany jako model nr 5 przez M. Aitkina i N. Longforda.

* Praca naukowa finansowana ze środków na naukę w latach 2010-2012, jako projekt badawczy nr N N111 194439.

Dane, które wykorzystano w części empirycznej, dotyczą uczniów, którzy egzamin maturalny zdawali w 2010 r. Prześledzono ich wyniki w nauce, począwszy od szkoły podstawowej przez gimnazjum, kończąc na egzaminie maturalnym. Wyniki badań z jednej strony mogą być wskazówką dotyczącą lokalizacji szkół, a z drugiej wspomagać decyzje o zmianie miejsca zamieszkania.

1. Edukacyjna wartość dodana w modelu Aitkina-Longforda

W dalszej części pracy będą stosowane następujące oznaczenia:

- x_{ij} − liczba punktów wejściowych uzyskanych przez i tego ucznia, którego miejsce kończenia szkoły (gimnazjum lub liceum) jest opisane indeksem j,
- y_{ij} − liczba punktów wyjściowych uzyskanych przez i tego ucznia, którego miejsce kończenia szkoły (gimnazjum lub liceum) jest opisane indeksem j,
- n_j − liczba uczniów w j-tej kategorii,
- k − liczba analizowanych kategorii ($k = 4$),
- n − liczba wszystkich uczniów, tzn. $n = n_1 + \ldots + n_k$,
- j − liczba indeksująca kategorie $j \in \{1,2,3,4\}$,
- $\overline{x}_j$, $\overline{y}_j$ − średni wynik odpowiednio wejściowy oraz wyjściowy na poziomie j-tej kategorii.

Dane wykorzystywane w artykule są nazywane niezbilansowanymi danymi panelowymi. Model, który zastosowano to model z czynnikami losowymi. W ekonomii model ten zawdzięcza popularność dzięki artykułowi Balestry i Nerlove'a [1966] mówiącym o popycie na gaz ziemny. Model ten ma postać:

$$y_{ij} = \alpha + \beta x_{ij} + \xi_j + e_{ij}. \tag{1}$$

W tym przypadku zakłada się że :

- e_{ij} jest zmienną losową o rozkładzie $N(0,\sigma^2)$,
- ξ_j jest zmienną losową o rozkładzie $N(0,\sigma_I^2)$,
- zmienne losowe e_{ij} dla różnych szkół i dla różnych uczniów są nieskorelowane,
- zmienna losowa ξ_j jest nieskorelowana z e_{ij} (tzn. $E(\xi_i, e_{ij}) = 0$).

Z powyższych założeń wynika, że:

$$\begin{aligned}\operatorname{var}(y_{ij}) &= \operatorname{var}(\xi_j + e_{ij}) = E(\xi_j + e_{ij})^2 - E^2(\xi_j + e_{ij}) \\ &= E(\xi_j^2 + 2\xi_j e_{ij} + e_{ij}^2) = \sigma_I^2 + \sigma^2,\end{aligned} \tag{2}$$

$$\mathrm{cov}(y_{ij}, y_{pj}) = \mathrm{cov}((\xi_j + e_{ij}), (\xi_j + e_{pj})) = E(\xi_j^2 + \xi_j e_{ij} + \xi_j e_{pj} + e_{ij} e_{pj}) = \sigma_I^2, \quad (3)$$

$$\rho = cor(y_{ij}, y_{pj}) = \frac{\sigma_I^2}{\sigma_I^2 + \sigma^2}. \quad (4)$$

Współczynniki tak określonego modelu szacuje się metodą największej wiarygodności, np. Aitkin i Longford [1986], lub uogólnioną metodą najmniejszych kwadratów, wspomnianą przez Baltaga (2005). Estymatory parametrów α oraz β mają postać:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (X^T V^{-1} X)^{-1} X^T V^{-1} y, \quad (5)$$

gdzie:

$$X^T = \begin{bmatrix} 1 & ,\dots, & 1 & ,\dots, & 1 & ,\dots, & 1 \\ x_{1,1} & ,\dots, & 1 & ,\dots, & x_{1,k} & ,\dots, & x_{k,n_k} \end{bmatrix},$$
$$y^T = \begin{bmatrix} y_{1,1} & ,\dots, & y_{n_1,} & ,\dots, & y_{k,1} & y_{k,n_k} \end{bmatrix}$$

oraz V jest macierzą kowariancji wektora y. Zgodnie z założeniami modelu macierz ta jest postaci:

$$V = \begin{bmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Omega_k \end{bmatrix},$$

gdzie:

$$\Omega_j = \begin{bmatrix} \sigma^2 + \sigma_I^2 & \sigma_I^2 & \dots & \sigma_I^2 \\ \sigma_I^2 & \sigma^2 + \sigma_I^2 & \cdots & \sigma_I^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_I^2 & \sigma_I^2 & \dots & \sigma^2 + \sigma_I^2 \end{bmatrix}_{n_j \times n_j}.$$

Po uproszczeniach wzoru (5) otrzymuje się*:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_{j=1}^{k} w_j & \sum_{j=1}^{k} w_j \bar{x}_j \\ \sum_{j=1}^{k} w_j \bar{x}_j & \sum_{j=1}^{k}\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^{k} w_j \bar{x}_j^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{k} w_j \bar{y}_j \\ \sum_{j=1}^{k}\sum_{j=1}^{n_j} (y_{ij} - \bar{y}_j)(x_{ij} - \bar{x}_j) + \sum_{j=1}^{k} w_j \bar{x}_j \bar{y}_j \end{bmatrix} \tag{6}$$

$$\hat{\beta} = \frac{\sum_{j=1}^{k}\sum_{j=1}^{n_j} (y_{ij} - \bar{y}_j)(x_{ij} - \bar{x}_j) + A_{xy}^*}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + A_{xx}^*}, \tag{7}$$

gdzie:

$$A_{xx}^* = \sum_{j=1}^{k} w_j (\bar{x}_j - \bar{x}^*)^2 \text{ oraz } A_{xy}^* = \sum_{j=1}^{k} w_j (\bar{x}_j - \bar{x}^*)(\bar{y}_j - \bar{y}^*) \ , \tag{8}$$

$$\bar{x}^* = \sum_{j=1}^{k} w_j \bar{x}_j \Big/ \sum_{j=1}^{k} w_j \text{ oraz } \bar{y}^* = \sum_{j=1}^{k} w_j \bar{y}_j \Big/ \sum_{j=1}^{k} w_j \tag{9}$$

oraz

$$w_j = n_j \sigma^2 / (\sigma^2 + n_j \sigma_I^2) . \tag{10}$$

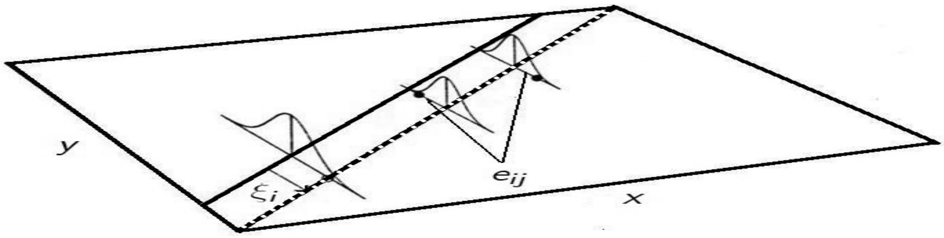
$$\operatorname{var}(\hat{\beta}) = \frac{\sigma^2 + \sigma_I^2}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + A_{xx}^*} \tag{11}$$

Zestawienie szkół odbywa się za pomocą porównania wartości oczekiwanej zmiennej ξ_j (wzór (1)). Zmienna ta mówi, o ile od uśrednionego wyniku całej populacji odchyla się uśredniony wynik j-tej szkoły. Na rysunku 1 przerywaną linią został oznaczony uśredniony wynik j-tej szkoły, zaś ciągła linia przedstawia uśredniony wynik całej populacji (czynnik e_{ij} odpowiada za odchylenie od

---

* Zastosowanie w pracach Aitkina i Longforda [1986].

uśrednianego wyniku na poziomie j-tej szkoły). Jeżeli wartość ξ_j jest dodatnia, to można powiedzieć, że j-ta szkoła poczyniła postęp w stosunku do uśrednionego wyniku całej populacji, jeśli zaś jest ujemna, to szkoła ta uzyskała wynik niższy niż uśredniony wynik badanej populacji.



Rys. 1. Schemat przedstawiający ideę pomiaru przyrostu wiedzy modelem Aitkina-Longforda

Źródło: [Skrondal, Rabe-Hesketh, 2008].

Chcąc oszacowywać wartość zmiennej losowej ξ_j (nie jest ona znana), wykorzystuje się twierdzenie o błędzie średniokwadratowym, według Jakubowskiego i Sztencela [2004]. W związku z tym, że wariancje σ^2 oraz σ_I^2 są znane przed oszacowaniem modelu, można tę informację wykorzystać jako informację a priori. Więcej na temat można przeczytać w pracy [Ejsmont, 2009], gdzie w szczegółach jest opisany cały algorytm estymacji, w tym również komponentów wariancji σ^2 i σ_I^2.

Wyznaczymy rozkład warunkowej zmiennej losowej ξ_j pod warunkiem $\bar{y}_j$ (podejście Bayesowskie). Ze wzoru (1) średnia na poziomie j-tej szkoły wyraża się wzorem:

$$\bar{y}_j = \alpha + \beta\bar{x}_j + \xi_j + \bar{e}_j. \tag{12}$$

Przy poczynionych założeniach, $\bar{y}_j$ ma rozkład normalny $N(\alpha + \beta\bar{x}_j, \sigma_I^2 + \sigma^2 / n_j)$. Ten rozkład można traktować jako rozkład a priori, a ponieważ ξ_j jest zmienną losową z rozkładu $N(0, \sigma_I^2)$, więc rozkład warunkowy $E(\xi_j \mid \bar{y}_j)$ też będzie rozkładem normalnym.

Znany jest fakt z rachunku prawdopodobieństwa, że jeżeli $X_1 \sim N(\mu_1, \sigma_1^2)$ i $X_2 \sim N(\mu_2, \sigma_2^2)$ oraz $\rho_{1,2} = cor(X_1, X_2)$, to rozkład warunkowy X_1/X_2 ma postać:

$$N\left(\mu_1 + \rho_{1,2}\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2), \sigma_1^2(1-\rho_{1,2}^2)\right). \tag{13}$$

Stąd uwzględniając to, że $\rho' = cor(\xi_j, \bar{y}_j) = \sigma_I^2 /(\sigma_I \sqrt{\sigma_I^2 + \sigma^2 / n_j})$, otrzymuje się, iż $E(\xi_j \mid \bar{y}_j)$ ma rozkład normalny:

$$N\left(\rho' \frac{\sigma_I}{\sqrt{\sigma_I^2 + \sigma^2 / n_j}}(\bar{y}'_j - \alpha - \beta \bar{x}_j), \sigma_I^2 (1-\rho'^2)\right)$$

lub w innym zapisie:

$$N(\rho n_j^* (\bar{y}_j - \hat{\alpha} - \hat{\beta}\bar{x}_j), n_j^*(1-\rho)\sigma_I^2 / n_j) \ , \tag{14}$$

gdzie:
$n_j^* = w_j /(1-\rho)$.

Przy porównaniu szkół wykorzystuje się wartość średnią z rozkładu warunkowego zadanego wzorem (14). Stąd efektywność nauczania, czyli edukacyjna wartość dodana ma postać:

$$e_j = \hat{\rho} n_j^* (\bar{y}_j - \hat{\alpha} - \hat{\beta}\bar{x}_j). \tag{15}$$

W celu sprawdzenia, czy uzyskane efekty losowe są istotne użyto testu Breuscha-Pagana [Baltagi, 2005; Hasio, 1999]. Jest to test mnożników Lagrange'a, w którym hipotezy są następujące: H_0 : $\sigma_I^2 = 0$, i alternatywna: $\sigma_I^2 \neq 0$.

Statystyka testowa ma postać:

$$LM = \frac{\left(\sum_{j=1}^{k} n_j\right)^2}{2\sum_{j=1}^{k} n_j(n_j - 1)} \left[\frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{n_j} e'_{ij}\right)^2}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} e'^2_{ij}} - 1\right]^2 \sim \chi^2(1) \ , \tag{16}$$

gdzie e'_{ij} są to reszty otrzymane w wyniku zastosowania metody MNK do wszystkich danych, niezależnie od szkół. Powyższy wzór mówi, że statystyka testowa *LM* przy założeniu hipotezy zerowej ma asymptotyczny rozkład chi-kwadrat z jednym stopniem swobody. Hipotezę zerową odrzuca się, jeżeli wartość statystyki *LM* należy do prawostronnego obszaru krytycznego.

2. Opis danych i uzyskane wyniki

Analizowane dane opisują wyniki edukacyjne uczniów kończących polskie licea w 2010 r. W każdym z poniższych typów zdawanych egzaminów przyjęto tę samą konwencję odnośnie skalowania danych. Zarówno punkty gimnazjalne, jak i maturalne zostały przeskalowane do poziomu 100, tzn. wynik danego ucznia, gimnazjalny i maturalny, podzielono przez maksymalną liczbę punktów możliwych do zdobycia oraz pomnożono przez 100. Pozwala to na łatwą interpretacje otrzymanych wyników. Bez trudu można zauważyć, czy uczeń polepszył, czy pogorszył swój wynik (%).

Dane, jakie przeanalizowano, reprezentują cztery kategorie możliwych miejsc kończenia szkoły gimnazjalnej lub licealnej:

- wieś − oznaczono przez 1,
- miasto do 20. tys. mieszkańców − oznaczono przez 2,
- miasto od 20 do 100 tys. mieszkańców − oznaczono przez 3,
- miasto powyżej 100 tys. mieszkańców − oznaczono przez 4.

W tabeli 1 zaprezentowano średnie wyniki w zależności od opisanych różnych kategorii możliwości kończenia szkoły. Warto podkreślić, że występowanie liceów na wsi jest rzadkością, aczkolwiek w każdym województwie takie licea się znajdują. Z tego powodu ta kategoria uczniów jest zdecydowanie najmniej licznie reprezentowana.

Tabela 1

Średnie wyniki różnych typów egzaminów przedstawione dla uczniów zdających maturę w liceum w 2010 r.

	Charakterystyki uczniów w momencie kończenia gimnazjum				Charakterystyki uczniów w momencie kończenia liceum				
Kategoria	liczba uczniów	średnia P	średnia G-H	średnia G-MP	liczba uczniów	średnia G-H	średnia G-MP	średnia M-P	średnia M-M
1	53766	74,46	75,37	61,21	4167	67,79	50,82	55,90	55,40
2	38007	74,21	74,25	59,89	42010	72,98	57,66	61,09	63,43
3	47339	76,16	75,41	61,96	72106	75,54	61,85	63,65	67,60
4	55914	79,11	77,56	65,58	76743	78,00	66,11	65,40	70,30
Razem	195026	76,16	75,79	62,39	195026	75,79	62,39	63,62	67,51

Nota:

- średnia P − średni wynik testu szóstoklasisty,
- średnia G-H − średni wynik części humanistycznej egzaminu gimnazjalnego,
- średnia G-MP − średni wynik części matematyczno-przyrodniczej egzaminu gimnazjalnego,
- średnia M-P − średni wynik egzaminu maturalnego z języka polskiego (część podstawowa),
- średnia M-M − średni wynik egzaminu maturalnego z matematyki (część podstawowa).

Źródło: [Centralna Komisja Egzaminacyjna, 2010].

Analizowano dwa typy modeli. Pierwszy dotyczył przyrostu wiedzy humanistycznej obliczonej na podstawie odpowiednich części humanistycznych to znaczy wiedzy w gimnazjum na podstawie testu szóstoklasisty oraz części humanistycznej egzaminu gimnazjalnego. Analogicznie obliczono przyrost wiedzy dla przedmiotów ścisłych, tym razem obliczając w odpowiednich miejscach względem przedmiotów ścisłych. W wyniku tych obserwacji otrzymano cztery modele zaprezentowane w tabeli 2. Wariancja zmiennej losowej określającej przyrost wiedzy pojedynczych uczniów jest znacząco większa od wariancji opisującej rozproszenie międzyszkolne. Jest to powodem wzięcia do analizy tylko czterech obiektów. Zauważalny jest również znacząco „szybszy wzrost wiedzy" dla przedmiotów ścisłych (współczynniki beta). Oszacowane wartości testu *LM* wskazują na to, że σ_I^2 jest statystycznie istotny na poziomie istotności 0,01. Zasadne jest więc stosowanie modelu efektów losowych (związanych z σ_I^2).

Tabela 2

Podstawowe charakterystyki statystyczne modelu efektów losowych

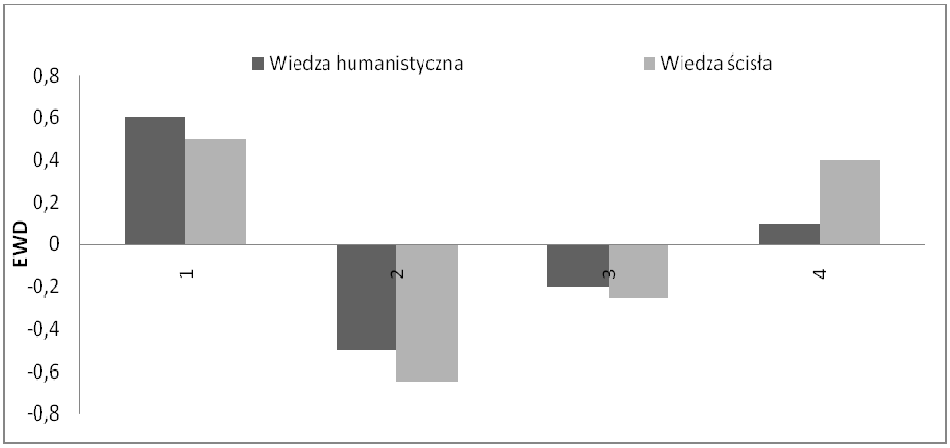
	Gimnazjum		Liceum	
	część humanistyczna	część ścisła	część humanistyczna	część ścisła
Wariancja σ^2	116,1536	208,185	173,8484	210,683
Wariancja σ_I^2	0,3241	0,3965	0,1545	1,336
Współczynnik beta	0,5365799	0,91571	0,5715188	0,7549003
Współczynnik alfa	34,8769272	-7,4213	19,580573	19,5840278
LM − p-value	< 0,01	< 0,01	< 0,01	< 0,01

Źródło: Obliczenia własne za pomocą programu Excel oraz R-project.

Podsumowanie

Rysunki 2 oraz 3 przedstawiają efektywność nauczania w zależności od miejsca położenia szkoły kończącej dany etap edukacji. Nie ma tutaj natomiast znaczenia, gdzie uczeń pobierał naukę we wcześniejszych etapach kształcenia.

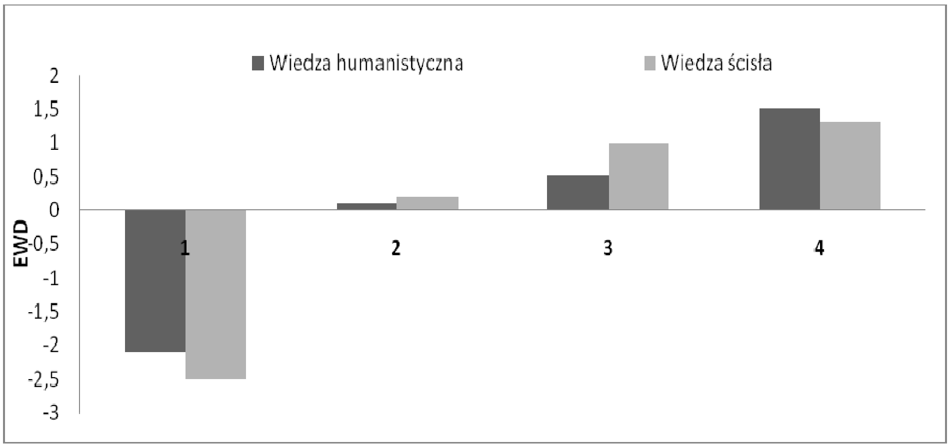
Interpretacja wyników przedstawionych na rysunku 2 nie jest trudna. Lokalizacja gimnazjum na terenie wiejskim nie wpływa negatywnie na wyniki uczniów. Co więcej, uczniowie takich placówek sumarycznie osiągają lepsze wyniki niż ich rówieśnicy pobierający naukę w większych ośrodkach. Można to tłumaczyć tym, że ewentualne braki w doborze wykwalifikowanej kadry dydaktycznej są rekompensowane mniejszym negatywnym oddziaływaniem środowiska. Zakładając nawet, że w ośrodkach wiejskich selekcja zawodowa jest ograniczona, samo przygotowanie nauczyciela na tym etapie kształcenia nie odgrywa kluczowej roli. Podobnie jeśli chodzi o dostęp do bibliotek czy innych ośrodków kultury, np. teatru, filharmonii. Rekompensowane jest to większym zaangażowaniem dzieci w proces nauczania. Negatywne wzorce, takie jak choćby zastępowanie czytania lektur przez czytanie ich skróconych opracowań, nie oddziałują tak silnie na to środowisko. Oprócz tego wydaje się, że sam proces wychowawczy w ośrodkach wiejskich jest zdecydowanie łatwiejszy. Nie słyszy się np., jak w przypadku większych ośrodków, o sporach uczniów i ich rodziców z nauczycielami. Na tym etapie kształcenia samo zaangażowanie ucznia, jego pilność i zdyscyplinowanie jest w stanie zrównoważyć gorszą infrastrukturę edukacyjną.



Rys. 2. Edukacyjna wartość dodana obliczona dla gimnazjów

Źródło: Na podstawie wyników z tabeli 1.

Do zdecydowanie innych wniosków prowadzi analiza wyników przedstawionych na rysunku 3. Widać wyraźnie, że osiągnięcia edukacyjne uczniów liceów ogólnokształcących zlokalizowanych na wsi istotnie odstają od osiągnięć ich rówieśników pobierających naukę w większych ośrodkach. Wniosku tego nie zmienia ewentualne uwzględnienie niewielkiej liczby takich placówek w porównaniu z liczbą wszystkich liceów w Polsce. Obraz sytuacji pozostaje bowiem taki sam po połączeniu pierwszych dwóch z rozpatrywanych kategorii lokalizacji, czyli tworzy się wspólna grupa ośrodków mniejszych niż dwudziestotysięczne. Wyraźnie gorsze na tym etapie wyniki edukacyjne mniejszych ośrodków dowodzą słuszności poglądów dotyczących m.in. likwidacji zamiejscowych ośrodków dydaktycznych wyższych uczelni. Wraz ze wzrostem szczebla edukacji wzrasta również konieczność zapewnienia odpowiedniej kadry dydaktycznej i całego zaplecza edukacyjnego. Takie możliwości stwarzają na razie jedynie duże ośrodki miejskie.



Rys. 3. Edukacyjna wartość dodana obliczona dla liceów

Źródło: Na podstawie wyników z tabeli 1.

Literatura

Aitkin M., Longford N. (1986): Statistical Modelling Issues in School Effectiveness Studies. "Journal of the Royal Statistical Society", Vol. 149, No. 1, s. 1-43.

Balestra P., Nerlove M. (1966): Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. "Econometrica", Vol. 34, No. 3, s. 585-612.

Baltagi B. (2005): Econometric Analysis of Panel Data. John Wiley & Sons, New York.

Centralna Komisja Egzaminacyjna (2010).

Ejsmont W. (2009): Efektywność nauczania we wrocławskich liceach. Didactics of Mathematics 5-6 (9-10), Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław 2009, s. 79-88.

Hasio C. (1999): Analysis of Panel Data. Cambridge University Press, Cambridge.

Jakubowski J., Sztencel R. (2004): Wstęp do teorii prawdopodobieństwa. Script, Warszawa.

Skrondal A., Rabe-Hesketh S. (2008): Multilevel and Longitudinal Modeling Using Stata. Stata Press Publication – StataCorp LP, College Station, Texas.

THE IMPACT OF SCHOOL LOCATION ON STUDENT ACHIEVEMENT

Summary

The modern socio-economic situation and, in particular, migrations highlight the issue of training quality depending on the location of the school. It happens that students who change their place of stay change the environment in which they learn. These changes may affect the results of training measured by national tests. The content of the article is an analysis of these effects in connection with the location of the school.