

Mirośława Żurek

Analiza bezrobocia w powiatach przy użyciu modelu równań strukturalnych

Słowa kluczowe: *modele równań strukturalnych, SEM, bezrobocie w powiatach, zmienne niemierzalne*

Abstrakt: Celem artykułu jest analiza oraz ustalenie determinantów regionalnego zróżnicowania poziomu bezrobocia w powiatach w Polsce. Na podstawie posiadanej wiedzy teoretycznej oraz publikowanych wyników badań w tym zakresie analizie w niniejszym opracowaniu poddano zależności między wysokością stopy bezrobocia występującą w danym regionie a istniejącym w nim poziomem infrastruktury transportowej, poziomem wykształcenia, gęstością zaludnienia, liczbą miejsc pracy, odsetkiem osób zamieszkałych w mieście oraz pracujących w rolnictwie. W badaniu posłużono się metodologią modelowania równań strukturalnych (SEM) pozwalającą na jednoczesne uwzględnianie relacji pomiędzy niezależnymi i zależnymi oraz mierzalnymi (obserwowalnymi) jak i nieobserwowalnymi (ukrytymi) zmiennymi. Za zmienną niemierzalną w badaniu przyjęto infrastrukturę transportu, czyli możliwości dojazdu występujące w danym powiecie. Na podstawie przeprowadzonej analizy udało się wykazać istnienie związków pomiędzy wysokością stopy bezrobocia a istniejącym w powiecie poziomem wykształcenia mieszkańców, liczbą zawartych małżeństw, odsetkiem osób zamieszkałych w mieście oraz osób zatrudnionych w rolnictwie. Poprawnie też udało się uwzględnić w modelu zmienną niemierzalną, jaką są możliwości dojazdu, określając ją za pomocą odsetka ludności zamieszkałej w mieście, długości dróg utwardzonych w powiecie, jak i długości tras komunikacyjnych. Otrzymane wyniki pozwalają na dogłębną i szeroką analizę w zakresie poznania przyczyn zróżnicowania poziomu bezrobocia w poszczególnych powiatach całego kraju.

WPROWADZENIE

Celem artykułu jest przeprowadzenie analizy regionalnego zróżnicowania poziomu bezrobocia w Polsce w powiatach oraz ustalenie jego determinantów. Na podstawie posiadanej wiedzy teoretycznej oraz publikowanych wyników badań w tym zakresie (por. Borowski, 2004; Śleszyński, 2007) analizie w niniejszym opracowaniu poddano zależności między wysokością stopy bezrobocia

występującą w danym regionie a istniejącym w nim poziomem infrastruktury transportowej, poziomem wykształcenia, gęstością zaludnienia, liczbą miejsc pracy, odsetkiem osób zamieszkałych w mieście oraz pracujących w rolnictwie. Dla celów analizy posłużono się modelem równań strukturalnych, w którym uwzględniono zarówno czynniki występujące po stronie podaży na rynku pracy (poziom wykształcenia, możliwości dojazdu do pracy), jak i po stronie popytu (istniejące miejsca pracy), a mające wpływ na poziom bezrobocia w danym powiecie. W badaniu ze względu na trudności pomiarowe bezrobocia ukrytego uwzględniono tylko poziom bezrobocia rejestrowanego podawany przez Główny Urząd Statystyczny. Dane dotyczące wysokości stopy bezrobocia liczonej według Badania Aktywności Ekonomicznej Ludności (BAEL) nie są dostępne na poziomie powiatów.

W dotychczasowym dorobku naukowym polskich badaczy niewiele jest opracowań zajmujących się statystyczną bądź ekonometryczną analizą regionalnego zróżnicowania wysokości stopy bezrobocia na poziomie powiatów, które obejmowałyby cały kraj. Badanie takie, jednakże tylko dla województwa lubelskiego, przeprowadził Borowski (2004). Stosując metodę grupowania struktur, wykazał on zależność pomiędzy wysokością stopy bezrobocia w danym powiecie a wiekiem i poziomem wykształcenia bezrobotnych. Analizę statystyczną na poziomie gmin przeprowadził m.in. Śleszyński (2007), wskazując, iż wysokość stopy bezrobocia w danej gminie jest odwrotnie proporcjonalna do gęstości zaludnienia oraz że położenie gmin wzdłuż ważnych szlaków komunikacyjnych i lepsze wskaźniki wykształcenia mieszkańców współwystępują z niższym poziomem bezrobocia. W tym samym przekroju badania zróżnicowania stóp bezrobocia prowadził również Matusik (2008), stosując w tym celu modele tendencji rozwojowej.

Większość analiz dotyczących regionalnego zróżnicowania wysokości stopy bezrobocia w Polsce dotyczyła dotąd przede wszystkim badań na poziomie województw. Analizy takie prowadził m.in. Tokarski (2005), który uzależnił stopę bezrobocia od wielkości produkcji w danym województwie, odwołując się do modelu wzrostu Harroda–Domara i keynesistowskich modeli rynków pracy. Pozostałe badania opierały się przede wszystkim na szeregach czasowych bezrobocia (por. Rozpędowska-Motraszek, 2006).

Żadne z wymienionych badań nie uwzględniło jednak zarówno popytowej, jak i podażowej strony rynku pracy przy badaniu stopnia zróżnicowania regionalnego stopy bezrobocia.

MODELE RÓWNAŃ STRUKTURALNYCH – SEM

Złożona struktura zjawisk ekonomicznych wymaga stosowania coraz bardziej wyszukanych modeli, które mogłyby precyzyjnie je odzwierciedlać. Konieczne

staje się włączenie do analizy zmiennych ukrytych (nieobserwowalnych). Na ich ujęcie w modelu ekonometrycznym pozwala metodologia modelowania równań strukturalnych (*Structural Equation Modeling* – SEM). Jest ona efektem połączenia analizy czynnikowej rozwijanej historycznie głównie w psychologii, jak i modelowania równań o charakterze przyczynowo-skutkowym stosowanego przez ekonometrię.

Modele równań strukturalnych definiuje się jako zestaw procedur i narzędzi statystycznych służących do pomiaru zależności przyczynowych w empirycznych badaniach naukowych. Metodologia SEM pozwala na jednoczesne uwzględnianie relacji między niezależnymi i zależnymi oraz mierzalnymi (obserwowalnymi) jak i nieobserwowalnymi (ukrytymi) zmiennymi. Ponadto możliwe jest ujmowanie w modelu potencjalnych błędów pomiaru wszystkich zmiennych obserwowalnych oraz estymowanie i testowanie wariancji oraz kowariancji między zmiennymi, a także badanie zależności bezpośrednich i pośrednich występujących pomiędzy nimi. W najprostszym ujęciu są to modele analizy ścieżkowej, w których dopuszcza się występowanie zmiennych ukrytych (por. Pearl, 2000).

Zmienne ukryte (niemierzalne) są teoretycznymi lub hipotetycznymi konstrukcjami bądź też zmiennymi o nieobserwowalnych realizacjach w danej próbie. W związku z tym nie istnieje bezpośrednia ani precyzyjna metoda ich pomiaru. Można je jedynie zmierzyć pośrednio, poprzez inne skorelowane z nimi zmienne.

W modelach SEM w związku ze sposobem konstruowania zmiennych ukrytych wyróżnia się podejście refleksyjne oraz formacyjne. W pierwszym przypadku zakłada się, że zmienne ukryte rzeczywiście istnieją, tylko mierzone są za pomocą silnie skorelowanych zmiennych obserwowalnych. Użycie więcej niż jednej zmiennej do pomiaru pozwala wówczas zminimalizować błędy estymacji oraz uzyskać pełniejszy i bardziej realny obraz analizowanej zmiennej niemierzalnej. W podejściu formacyjnym zmienna ukryta w rzeczywistości istnieć nie musi, a jedynie jest ona wynikiem oddziaływania pewnej liczby zmiennych obserwowalnych (por. Kaplan, 2000; Loehlin, 1987).

Pełny model równań strukturalnych składa się z modelu opisującego powiązania pomiędzy zmiennymi ukrytymi – model wewnętrzny oraz z modelu pomiaru endogenicznych i egzogenicznych zmiennych nieobserwowalnych - model zewnętrzny (por. Kaplan, 2000; Bollen, 1989).

Model wewnętrzny ma postać:

$$\eta = B\eta + \Gamma\xi + \zeta \quad (1)$$

gdzie:

$\eta_{m \times 1}$ – wektor endogenicznych zmiennych ukrytych,

$\xi_{k \times 1}$ – wektor egzogenicznych zmiennych ukrytych,

B_{mxm} – macierz współczynników regresji przy zmiennych endogenicznych, określająca ich wzajemny wpływ na siebie,

Γ_{mxk} – macierz współczynników przy zmiennych egzogenicznych, ukazująca ich wpływ na zmienne endogeniczne,

ζ_{mx1} – wektor składników losowych.

Założenia dotyczące składnika losowego są następujące:

$$E(\zeta) = 0$$

$$\sum_{\zeta} = \sigma^2 I$$

$$|I - B| \neq 0$$

Model zewnętrzny jest dany jako:

$$y = \Pi_y \eta + \varepsilon \quad (2)$$

$$x = \Pi_x \xi + \delta \quad (3)$$

gdzie:

y_{px1} – wektor obserwowalnych zmiennych endogenicznych,

x_{qx1} – wektor obserwowalnych zmiennych egzogenicznych

Π_y, Π_x – macierze ładunków czynnikowych o wymiarach odpowiednio:

$\varepsilon_{px1}, \mu_{qx1}$ – wektory błędów pomiaru

Założenia dotyczące składnika losowego w modelu zewnętrznym są następujące:

$$E(\varepsilon) = E(\delta) = 0$$

$$Cov(\eta, \varepsilon) = Cov(\xi, \delta) = 0$$

Warunkiem dostatecznym identyfikowalności modelu SEM ze względu na parametry jest nierówność (por. Bollen, 1989):

$$t \leq \frac{(p+q)(p+q+1)}{2} \quad (4)$$

gdzie:

t – liczba nieznanymi parametrów modelu,

$p+q$ – liczba zmiennych obserwowalnych (endogenicznych, – p i egzogenicznych – q).

W niniejszym opracowaniu uwzględniono jedynie związki o charakterze liniowym między zmiennymi obserwowalnymi a ukrytymi. Sama jednak metodologia SEM zezwala na modelowanie również zależności o charakterze nieliniowym.

OCENA STOPNIA DOPASOWANIA MODELU

Stopień dopasowania modelu równań strukturalnych określa się najczęściej przez porównanie otrzymanego modelu z modelem nasyconym i niezależnym. W pierwszym z nich zakłada się, że wszystkie zmienne są ze sobą skorelowane, w modelu niezależnym zaś, że korelacja nie występuje pomiędzy żadną spośród par zmiennych (por. Loehlin, 1987; Kaplan, 2000).

W niniejszym opracowaniu ograniczono się do opisu kilku najważniejszych miar, za pomocą których zweryfikowano dostosowanie modelu do danych w części empirycznej artykułu. Wśród miar stopnia dopasowania modelu SEM wyróżnia się między innymi:

– miary porównujące estymowany model z modelem bazowym:

IFI (*Incremental Fix Index*)

$$IFI = \frac{T_b - T_h}{T_b - df_h} \quad (5)$$

gdzie:

T_h – statystyka chi-kwadrat estymowanego modelu,

df_h – stopnie swobody modelu estymowanego,

T_b – statystyka chi-kwadrat modelu bazowego (niezależnego).

Wartości wskaźnika IFI powinny zawierać się w przedziale $<0; 1>$, przy czym model uznaje się za dobrze dopasowany, jeśli wartość tego współczynnika jest większa od 0,95. W literaturze dotyczącej SEM jak i stosownym oprogramowaniu w celu zbadania stopnia dopasowania modelu stosuje się również wskaźniki TFI, RFI, NFI, CFI, których wyliczanie i interpretacja opierają się na podobnej zasadzie co IFI (por. Bollen, 1989);

– średniokwadratowy błąd aproksymacji:

RMSEA (*Root Mean Square Error of Approximation*) – W przeciwieństwie do opisywanej miary IFI, podczas obliczania wskaźnika RMSEA nie następuje porównywanie modelu estymowanego z modelem bazowym. Wskaźnik ten oblicza się wg wzoru:

$$RMSEA = \sqrt{\frac{T_h - df_h}{(N - 1)df_h}} \quad (6)$$

Oznaczenia są analogiczne jak w formule IFI (5). Im niższa wartość wyliczonego na podstawie modelu wskaźnika RMSEA tym lepszy stopień dopasowania modelu. Przyjmuje się, że dla wartości RMSEA mniejszej od 0,05 model jest bardzo dobrze dopasowany do danych, natomiast przypadek wartości wskaźnika większej od 0,10 oznacza już złe dopasowania modelu do danych;

– miary zawartości informacyjnej:

kryterium informacyjne AIC, BCC lub BIC (por. Bollen, 1989).

W przeciwieństwie do powszechnie stosowanego we wnioskowaniu statystycznym centralnego rozkładu chi-kwadrat, w przypadku modeli SEM stosuje się jego wersję niecentralną (por. Curran, 2002).

ANALIZA ZRÓŻNICOWANIA POZIOMU BEZROBOCIA W POWIATACH

Poziom bezrobocia w Polsce jest zróżnicowany na terenie całego kraju. Obok powiatów ze stosunkowo niską stopą bezrobocia wynoszącą zaledwie kilka punktów procentowych istnieją regiony, gdzie poziom bezrobocia rejestrowanego przekracza 20% i stanowi ważny problem nie tylko ekonomiczny, ale i społeczny. By takim regionalnemu zróżnicowaniu na rynku pracy przeciwdziałać, konieczne jest najpierw ustalenie przyczyn, które owe dysproporcje w powiatach powodują. W tym celu postawiono hipotezę i oszacowano model zgodnie z metodologią modelowania równań strukturalnych (SEM).

Stawiana w artykule hipoteza modelowa jest częściowo uzależniona od zakresu udostępnianych przez Główny Urząd Statystyczny danych. W związku z tym w badaniu zabrakło m.in. analizy zależności pomiędzy wysokością bezrobocia rejestrowanego w powiatach a poziomem wykształcenia ogółu mieszkańców. Uwzględniony mógł być tylko poziom wykształcenia wśród zarejestrowanych bezrobotnych. Z powodu braku dostępnych danych nie posłużono się także bardziej prawidłowym wskaźnikiem, jakim jest stopa bezrobocia wg BAEL. Utrudnieniem przy badaniu różnic w zakresie bezrobocia pomiędzy powiatami okazał się również stosunkowo długi okres między końcem danego roku kalendarzowego a momentem pojawienia się szczegółowych danych.

W badaniu wykorzystano dane statystyczne za rok 2007 pochodzące z Banku Danych Regionalnych Głównego Urzędu Statystycznego. W celu oszacowania modelu posłużono się programem SPSS Amos v. 16 w wersji Demo.

W modelu wykorzystano następujące zmienne:

1) endogeniczne zmienne obserwowalne:

– skupienie – odsetek ludności zamieszkałej w miastach, wyrażony w procentach,

– autobus – długość tras komunikacji miejskiej na 100 km², wyrażona w km,

– drogi – długość dróg powiatowych o nawierzchni twardej na 100 km², wyrażone w km,

– bezrobocie – stopa bezrobocia rejestrowanego w powiecie wyrażona w procentach (wg GUS),

– rolnictwo – odsetek ludności zatrudnionej w sektorze rolnictwa indywidualnego,

– małżeństwa – liczba zawartych małżeństw na 1000 ludności,

– ferty – oferty pracy przypadające na 1000 ludności powiatu, wyrażone w sztukach;

2) egzogeniczne zmienne obserwowalne:
 – bez wykształcenia – stosunek bezrobotnych z wykształceniem podstawowym bądź gimnazjalnym do ogólnej liczby pozostających bez pracy, wyrażony w procentach;

3) egzogeniczne zmienne nieobserwowalne
 – możliwości dojazdu – zmienna nieobserwowalna (ukryta), utożsamiana z łatwością i dostępnością przemieszczania się w powiecie.

Używając określonych powyżej zmiennych sformułowano następującą hipotezę modelową:

– model wewnętrzny (analiza czynnikowa) zapisany został jako:

$$Drogi = \beta_1 \text{MozliwosciDojazdu} + e_5$$

$$Skupienie = \beta_2 \text{MozliwosciDojazdu} + e_6$$

$$\text{Autobus} = 1 \text{MozliwosciDojazdu} + e_7$$

$$\text{Bezrobocie} = \alpha_1 \text{Oferty} + \alpha_2 \text{MozliwosciDojazdu} + \alpha_3 \text{Rolnictwo} \\ + \alpha_4 \text{Malzenstwa} + \alpha_5 \text{Bez_wykształcenia} + e_1$$

Model rekursywny (zewewnętrzny) zaś określono w sposób następujący:

$$\text{Oferty} = \gamma_1 \text{MozliwosciDojazdu} + e_2$$

$$\text{Rolnictwo} = \gamma_2 \text{Skupienie} + e_3$$

$$\text{Malzenstwa} = \gamma_3 \text{Skupienie} + e_4$$

W trzecim równaniu modelu wewnętrznego przy zmiennej nieobserwowalnej „Możliwości_dojazdu” ustalono stały parametr o wartości równej 1, co było podyktowane koniecznością spełnienia założeń, co do identyfikowalności modelu.

Model oszacowano metodą największej wiarygodności. Otrzymane w wyniku estymacji parametry zestawiono w tabeli.

Dla poziomu istotności $\alpha = 0,05$ wszystkie parametry tego modelu poza α_1 są istotne statystycznie. Współczynnik IFI jest większy od 0,95 i wynosi 0,964, RMSEA = 0,084, a wartość kryterium AIC jest stosunkowo niska w porównaniu z modelem niezależnym i zbliżona do wartości modelu nasyconego. Dla modelu nasyconego kryterium AIC = 72,00, dla niezależnego AIC = 1327,96, a dla modelu obliczonego wynosi 101,15. Otrzymane oceny stopnia dopasowania modelu należy uznać więc za zadowalające.

Na podstawie wyestymowanego modelu możliwe jest sformułowanie następujących wniosków.

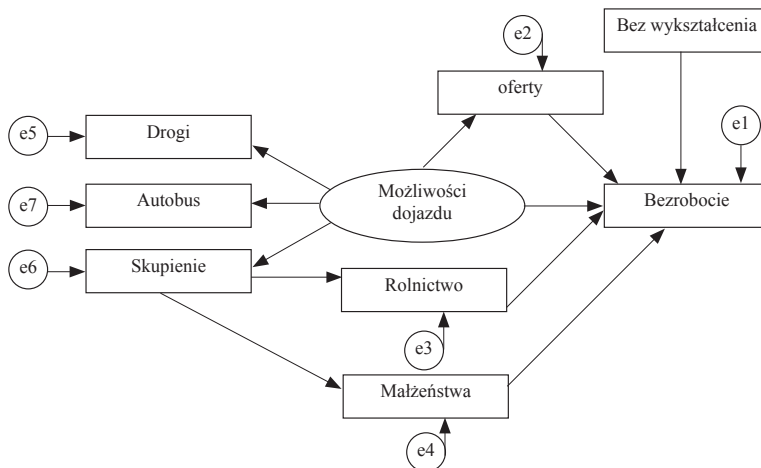
1. Poziom bezrobocie w powiecie jest uzależniony od liczby ofert pracy, poziomu wykształcenia ludności, liczby osób zatrudnionych w rolnictwie, a także liczby zawartych małżeństw przypadających na 1000 mieszkańców powiatu.

2. Im więcej osób o niskim poziomie wykształcenia, tym większe bezrobocie.

3. Im więcej ofert pracy w danym powiecie, tym poziom bezrobocia jest mniejszy, przy czym nowe oferty pracy zależą również od infrastruktury transportowej danego regionu. Jest to czynnik, na który przedsiębiorcy zwracają uwagę podczas lokalizacji swoich przedsiębiorstw, a więc i tworzenia nowych miejsc pracy. W uzyskanym modelu *Możliwości_dojazdu* są dodatnio skorelowane ze zmienną *Oferty*, co potwierdza, iż im lepsza infrastruktura transportowa w regionie, tym więcej nowych miejsc pracy.

Według wyników modelu w powiatach, w których więcej osób zatrudnionych jest w sektorze rolnictwa indywidualnego, większe jest też bezrobocie.

Rys. 1. Diagram modelu równań strukturalnych opisujący model wewnętrzny i zewnętrzny hipotezy modelowej dotyczącej determinantów zróżnicowania stopy bezrobocia w Polsce



Źródło: opracowanie własne.

Jednocześnie liczba osób pracujących w rolnictwie jest związana z liczbą osób zamieszkałych na wsi, stąd też, wg modelu, jeśli odsetek ludności zamieszkałej w mieście wzrośnie, to odsetek ludności pracującej indywidualnie w rolnictwie zmniejszy się.

Zgodnie z otrzymanymi wynikami w powiatach, gdzie zawarto więcej małżeństw w przeliczeniu na 1000 ludności, jest jednocześnie wyższe bezrobocie. Analiza wykazała także istnienie dodatniej korelacji pomiędzy odsetkiem osób zamieszkałych na wsi a liczbą zawieranych małżeństw.

Na definicję zmiennej określającej możliwości dojazdu składają się: długość dróg o nawierzchni utwardzonej, długość tras komunikacyjnych, a także odsetek ludności zamieszkałej w mieście. Im lepsze możliwości dojazdu do pracy w danym powiecie, tym niższe jest bezrobocie.

Dodatkowo w powiatach, gdzie odsetek ludności zamieszkałej w mieście jest większy, a więc występuje większe skupienie ludności, bezrobocie jest niższe.

Tabela. Wyniki estymacji modelu równań strukturalnych

			Obliczo- ny para- metr	błąd standar- dowy	statysty- ka t-stu- denta	Ozna- czenie
Skupienie	<--	Mozliwosci_dojazdu	,2591	,0149	17,3987	β_2
Rolnictwo	<--	Skupienie	-,0064	,0002	-28,4476	γ_2
Malzenst	<--	Skupienie	-,0049	,0012	-4,1463	γ_3
Oferty	<--	Mozliwosci_dojazdu	,0075	,0013	5,8135	γ_1
Autobus	<--	Mozliwosci_dojazdu	1,0000			
Drogi	<--	Mozliwosci_dojazdu	,2826	,0130	21,7932	β_1
Bezroboci	<--	Mozliwosci_dojazdu	-,0122	,0048	-2,5247	α_2
Bezroboci	<--	Bez_wyksztalcenia	0,3184	0,0465	6,8351	α_5
Bezroboci	<--	Rolnictwo	2,6505	1,3592	1,9500	α_3
Bezroboci	<--	Malzenstwa	1,5146	,4542	3,3349	α_4
Bezroboci	<--	Oferty	-,2861	,1602	-1,7860	α_1

Źródło: obliczanie własne na podstawie danych GUS.

ZAKOŃCZENIE

W badaniu wykazano istnienie związków między wysokością stopy bezrobocia w danym powiecie a istniejącym w nim poziomem wykształcenia mieszkańców, liczbą zawartych małżeństw, odsetkiem osób zamieszkałych w mieście oraz osób zatrudnionych w rolnictwie. Poprawnie też uwzględniono w modelu zmienną niemierzalną, jaką są możliwości dojazdu, określając ją za pomocą odsetka ludności zamieszkałej w mieście, długości dróg utwardzonych w powiecie oraz długości tras komunikacyjnych. Otrzymane wyniki pozwalają na szczegółową analizę przyczyn zróżnicowania poziomu bezrobocia w poszczególnych powiatach całego kraju.

LITERATURA

- Bollen, K. A. (1989), *Structural Equations with Latent Variables*, Wiley.
- Borowski P. (2004), *Badanie bezrobocia metodą grupowania struktury na przykładzie woj. Lubelskiego*, „Wiadomości Statystyczne”, nr 2.
- Curran P. J., Bollen K. A., Paxton P., Kirby J., Chen F. (2002), *The Noncentral Chi-square Distribution in Misspecified Structural Equation Models: Finite Sample Results from a Monte Carlo Simulation*, „Multivariate Behavioral Research”, vol. 37, No. 1,
- Kaplan D. (2000), *Structural Equation Modeling: Foundations and Extensions*, Sage Publications.
- Loehlin J. C. (1987), *Latent variable models: An introduction to factor, path, and structural analysis*, Erlbaum.

- Matusik St. (2008), *Kształtowanie się stopy bezrobocia w gminach woj. Małopolskiego*, „Wiadomości Statystyczne”, nr 1.
- Pearl J. (200), *Causality. Models, reasoning and inference*, Cambridge,
- Rozpędowska-Matraszek D. (2006), *Prognozowanie bezrobocia według województw*, „Wiadomości Statystyczne”, nr 12.
- Śleszyński P. (2007), *Zmiany liczby bezrobotnych w gminach*, „Wiadomości Statystyczne”, nr 2.
- Tokarski T., (2005), *Regionalne zróżnicowanie rynku pracy*, „Wiadomości Statystyczne”, nr 11.

USING STRUCTURAL EQUATION MODELING TO ANALYZE UNEMPLOYMENT IN DISTRICTS

Key words: *structural equation model, SEM, unemployment in districts, latent variables*

Abstract: The aim of this article is to analyze and discover reasons for district's unemployment rate variety in Poland. According to theory and published findings the study concentrates on relationships between unemployment rate in region and economic although infrastructure and social factors. The study was performed using structural equation modelling (SEM) in which relationships between dependent and independent although latent and measurable variables can be include. As a latent variable in this research transport infrastructure was adopted. It was defined using urbanization rate, road and communication way length. The research demonstrated the existence of non-positive relationships between unemployment rate in district and urbanization rate. Low educational level and high percentage of people employed in agriculture increase unemployment level in districts. Good transport infrastructure has positive influence on number of vacancy and unemployment rate. Results of the research allow to analyze in detail the reasons for unemployment rate variety in various districts.