

Danuta Rozpędowska-Matraszek, Karolina Lewandowska
Uniwersytet Łódzki

ZASTOSOWANIE METOD AGLOMERACYJNYCH I MODELOWANIA WEKTOROWO-AUTOREGRESYJNEGO W ANALIZIE PRZESTRZENNO-CZASOWEJ RYNKU PRACY W POLSCE

1. Wstęp

Modele wektorowo-autoregresyjne sformułowane przez C.A. Simsa [1980] są modelami wielorównaniowymi, które zyskują obecnie coraz bardziej ugruntowaną pozycję wśród metod ekonometrycznych. Pozwalają one badać zależności pomiędzy zmiennymi, testować ogólnie sformułowane teorie ekonomiczne, a przede wszystkim dają możliwość efektywnego prognozowania przyszłych wartości zmiennych [Kusideł 2000].

Modele wektorowo-autoregresyjne są zazwyczaj budowane na szeregach czasowych, jednakże ostatnio podejmowane są również próby zastosowań danych panelowych (próby przekrojowo-czasowe), które dają więcej informacji szczegółowych na temat systemów ekonomicznych.

Zasadniczym celem badań autorek było zbudowanie modelu VAR (ang. *vector autoregressive*) opisującego rynek pracy w Polsce według województw. W opracowaniu prezentowane są wyniki tych badań, będących jedną z pierwszych prób połączenia zagadnienia paneli niezbilansowanych z modelowaniem wektorowo-autoregresyjnym oraz wykorzystania metod aglomeracyjnych do redukcji wymiarów modelu.

2. Przegląd aktualnych badań

Problemy modelowania wektorowo-autoregresyjnego z wykorzystaniem prób przestrzenno-czasowych są coraz częściej dyskutowane w publikacjach zagranicznych i krajowych.

Gromadzenie danych makroekonomicznych w dużych panelach sprawiło, iż wśród ekonomistów wzrosło zainteresowanie estymacją dynamicznych modeli ekonometrycznych na danych panelowych. Zagadnienie to stało się jeszcze bardziej istotne w badaniach regionalnych, gdyż dzięki zastosowaniu danych panelowych do estymacji związków pomiędzy obiektami (np. krajami) uzyskujemy jednocześnie efekty specyficzne dla poszczególnych obiektów. Wiele makroekonomicznych paneli ma znacznie większy wymiar czasu (T) niż obiektów (N). Różnica ta jest bardzo istotna przy doborze techniki estymacji, którą należy zmieniać wraz ze zmianą wymiaru panelu. Tematyką estymacji dynamicznych modeli ekonometrycznych zajęły się R.A. Hudson i A.L. Owen [1996], które po przeprowadzeniu wielu badań opracowały przewodnik służący do wyboru odpowiedniej techniki estymacji dla paneli o różnych rozmiarach oraz badały problemy paneli niezbilansowanych. Autorki udowodniły, iż odpowiedni wybór estymatora sprawi, że otrzymamy poprawne oszacowania parametrów modelu.

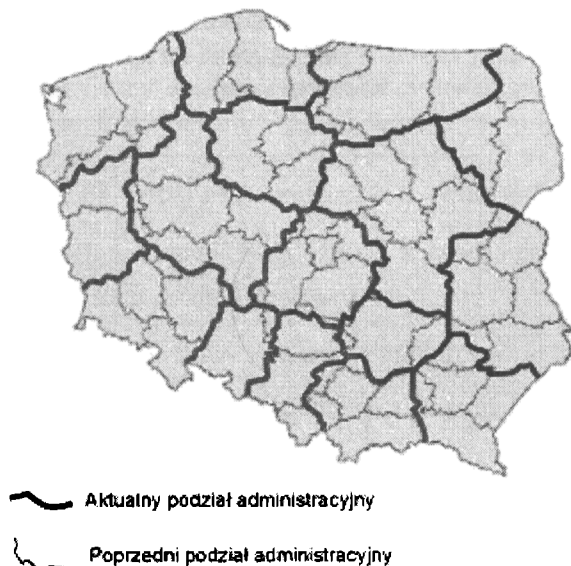
Modelowanie wektorowo-autoregresyjne na danych panelowych do zbadania związków pomiędzy sprzedażą i przepływem gotówki oraz inwestycjami i nakładami na badania i rozwój wykorzystali w swoich analizach również B.H. Hall, J. Mairesse, L. Branstetter, B. Crepon [1998]. W tym celu użyto danych przestrzenno-czasowych dotyczących stu dwudziestu firm wysokich technologii (ang. *high technology*) z Francji, Japonii i Stanów Zjednoczonych działających w latach 1978-1989. W badaniu zastosowano maksymalnie 5 opóźnień, a estymacji modelu dokonano za pomocą uogólnionej metody momentów GMM (ang. *generalized method of moments*).

Problem modelowania wektorowo-autoregresyjnego podjął również M. Arellano [2003], który opisał nowe metody estymacji dynamicznych modeli ekonometrycznych opartych na danych panelowych. Zaprezentował 3 metody, tj. GMM (która jest najbardziej efektywna dla krótkich paneli), PML (ang. *pseudo-maximum likelihood*) oraz SIV (ang. *semiparametric instrumental variable*). Jako empiryczny przykład posłużył mu skonstruowany przez niego samego model VAR dla dwóch zmiennych (zatrudnienia i płac przeciętnych) z jednym opóźnieniem, który w obydwu równaniach zawiera efekty indywidualne i czasowe.

3. Prezentacja modelu

Dane statystyczne zastosowane w analizach dotyczą stopy bezrobocia, inflacji, płac przeciętnych oraz wydajności pracy i obejmują okres od IV kw. 1996 r. do IV kw. 2004 r. W niniejszym artykule wykorzystano dane kwartalne o stopie bezrobocia i płacach przeciętnych w Polsce, według podziału administracyjnego obowiązującego od 1 stycznia 1999 r. Przeprowadzone zmiany administracyjne wymusiły konieczność przeliczenia danych z lat 1996-1998 na aktualnie obowiązujący podział kraju.

W celu przedłużenia próby przy użyciu programu komputerowego ARCVIEW¹ nałożono na siebie mapy Polski w obu podziałach administracyjnych (rys. 1).



Rys. 1. Udziały powierzchni
Źródło: opracowanie własne.

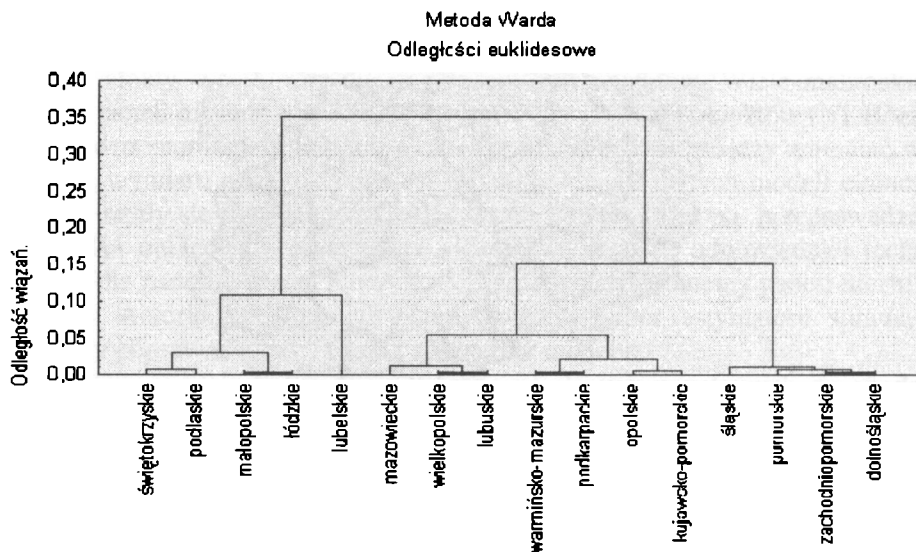
Dane o poziomie inflacji w województwach zaczęto zbierać w wojewódzkich urzędach statystycznych dopiero od I kw. 2001 r., dlatego w celu uzupełnienia brakujących danych została wykonana prognoza wsteczna za pomocą modeli szeregów czasowych z wahaniami okresowymi zmiennej prognozowanej. W rezultacie uzyskano kompletny panel danych przestrzenno-czasowych składający się z 528 obserwacji.

W celu zredukowania wymiaru danych panelowych zastosowano jedną z metod wielowymiarowej analizy porównawczej, tj. metodę aglomeracyjną, która pozwoliła wyodrębnić województwa najbardziej podobne do siebie pod względem średniego poziomu wszystkich badanych zmiennych w okresie od 1999-2004 r. Ze względu na to, iż zmienne są wyrażone w różnych jednostkach, należało doprowadzić je do wzajemnej porównywalności, dlatego też dokonano normalizacji zmiennych, stosując standaryzację [Malina 2004].

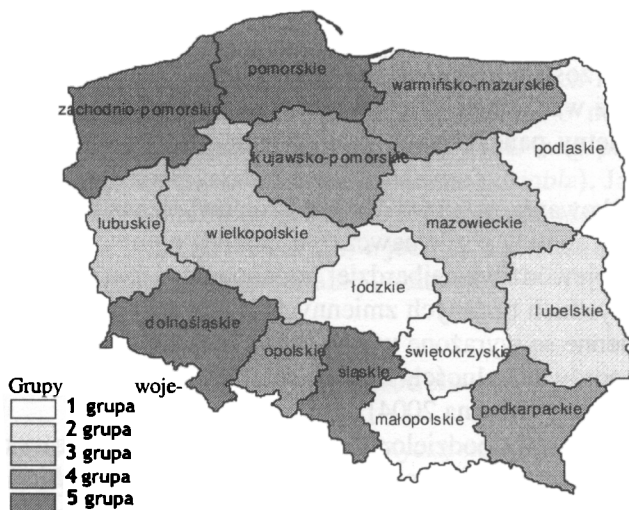
Województwa zostały podzielone na grupy (zwane również skupieniami) tak, że obiekty znajdujące się w jednej grupie były podobne do siebie i różne od pozo-

¹ Podstawę podziału stanowiła powierzchnia. Program umożliwił wyznaczenie procentowych udziałów powierzchni starych województw w nowych. Do obliczeń przyjęto założenie o równomierności rozłożenia cech w przestrzeni (co nie jest zgodne z rzeczywistością gospodarczą, a przez to zmniejsza dokładność otrzymanych wyników). Opisanego przeliczenia dokonano z powodu braku bardziej adekwatnych metod.

stałych grup. Do analiz zastosowano metodę minimalnych wariancji (Warda), która umożliwiła zredukowanie wymiaru danych panelowych z szesnastu obiektów do pięciu grup województw, co przedstawia rys. 2.



Rys. 2. Podział obiektów na grupy
Źródło: obliczenia własne w pakiecie Statistica.



Rys. 3. Grupy województw zastosowane w analizach
Źródło: opracowanie własne w programie ARCVIEW.

Na mapie na rys. 3 możemy zaobserwować układ przestrzenny uzyskanych grup województw. Jak widać, nie pokrywają się one z geograficznymi regionami

Polski, jednakże odwzorowują jednostki podobne pod względem poziomu rozwoju gospodarczego.

Na podstawie powyższego podziału otrzymujemy tzw. panel niezbilansowany, tj. panel danych statystycznych składający się z pięciu grup województw, w których znajdują się różne liczby obiektów, co stwarza dodatkowe problemy przy estymacji modelu ekonometrycznego.

3.1. Model panelowy niezbilansowany – model statyczny

Korzystanie z danych panelowych umożliwia rozwiązanie problemów ekonomicznych, których nie można zweryfikować empirycznie na podstawie odrębnych analiz czasowych lub przekrojowych. Wykorzystanie danych panelowych do estymacji daje możliwość jednoczesnego wykorzystania informacji zarówno o dynamice czasowej, jak i o indywidualnej specyfice obserwowanych obiektów. Dlatego kolejnym etapem rozważań na temat rynku pracy w Polsce było dokonanie oszacowań parametrów modelu panelowego niezbilansowanego dla zmiennych w postaci:

$$\begin{aligned} \ln U_{it} &= \alpha_0 + \alpha_{0i} + \alpha_1 \ln(U_{-1t}) + \alpha_2 \ln(W_{-1t}) + \alpha_3 \ln(V_{-1t}) + \alpha_4 \ln(PC_{-1t}) + \dots + \alpha_8 \ln(PC_{-2t}) + \xi_{it} \\ \ln W_{it} &= \alpha_0 + \alpha_{0i} + \alpha_1 \ln(U_{-1t}) + \alpha_2 \ln(W_{-1t}) + \alpha_3 \ln(V_{-1t}) + \alpha_4 \ln(PC_{-1t}) + \dots + \alpha_8 \ln(PC_{-2t}) + \xi_{it} \\ \ln V_{it} &= \alpha_0 + \alpha_{0i} + \alpha_1 \ln(U_{-1t}) + \alpha_2 \ln(W_{-1t}) + \alpha_3 \ln(V_{-1t}) + \alpha_4 \ln(PC_{-1t}) + \dots + \alpha_8 \ln(PC_{-2t}) + \xi_{it} \end{aligned} \quad (1)$$

gdzie: $i = 1, \dots, 5$ grup województw; $t = 1, \dots, 33$ okresów ($t = 1996.4, \dots, 2004.4$); U – stopa bezrobocia, PC – inflacja, W – płace przeciętne, V – wydajność pracy.

W wyniku estymacji powyższych równań uzyskano tzw. model FEM² z efektami grupowymi oraz model REM³ z dekompozycją składnika losowego, które umożliwiły analizę rynku pracy w Polsce w podziale na pięć grup województw. Ze względu na to, iż niektóre z parametrów modelu były nieistotne statystycznie (przyjmując krytyczną wartość statystyki t -Studenta na poziomie istotności 0.05), usuwano je w sposób sekwencyjny, dlatego uzyskano oszacowania parametrów przedstawione w tab. 1.

Uzyskane oszacowania parametrów strukturalnych w obu modelach są zbliżone oraz zgodne z założeniami teoretycznymi ekonomii. Różnice między nimi są statystycznie istotne, co potwierdził zarówno test współczynnika wiarygodności, jak i test F (test Chowa). Z kolei wartości statystyk t -Studenta świadczą o istotności wpływu każdego z nich na zmienną objaśnianą.

Na podstawie uzyskanych rezultatów można sformułować następujące wnioski.

1. Stopa bezrobocia jest uzależniona w krótkim okresie od opóźnionych o jeden kwartał stopy bezrobocia, wynagrodzeń i inflacji, w długim zaś okresie od opóźnionych o dwa kwartały: stopy bezrobocia i wydajności pracy.

² Model ze zmiennymi sztucznymi, w którym efekty grupowe traktuje się jako część składową wyrazu wolnego.

³ Efekty grupowe traktowane są jako składowa składnika losowego.

Tabela 1. Oszacowania parametrów modeli

Równanie	Zmienna	Model ze zmiennymi sztucznymi		Model z dekompozycją składnika losowego	
Stopa bezrobocia (U_t)	$\ln(U_{1t})$	1,05302	(28,14)	1,06175	28,43
	$\ln(U_{2t})$	-0,11775	(-3,26)	-0,12576	-3,49
	$\ln(W_{1t})$	-0,11535	(7,49)	-0,09969	6,69
	$\ln(V_{2t})$	-0,18779	(-9,80)	-0,15908	-8,95
	$\ln(Pc_{1t})$	-2,06880	(-12,73)	-2,07410	-12,82
	wyraz wolny			10,50484	13,49
	R^2	0,976		0,967	
	var (e_{it})			0,002548	
	var (u_i)			0,000408	
Wynagrodzenia przeciętne (W_t)	$\ln(U_{1t})$	0,51542	11,38	0,51155	11,34
	$\ln(U_{2t})$	-0,55893	-12,74	-0,55764	-12,71
	$\ln(W_{2t})$	0,78906	43,82	0,81519	53,56
	$\ln(V_{1t})$	-0,14435	5,75	0,09434	5,37
	wyraz wolny			0,66648	5,85
	R^2	0,952		0,951	
	var (e_{it})			0,004156	
	var (u_i)			0,000053	
Wydajność pracy (V_t)	$\ln(W_{1t})$	0,09421	5,40	0,06849	4,26
	$\ln(V_{1t})$	0,41654	13,34	0,45453	15,38
	$\ln(V_{2t})$	0,41435	14,38	0,43631	15,45
	$\ln(Pc_{1t})$	-3,92593	-19,96	-3,90711	-19,88
	$\ln(Pc_{2t})$	2,93846	13,50	3,08256	14,39
	wyraz wolny			4,33709	3,26
	R^2	0,956		0,951	
	var (e_{it})			0,003393	
	var (u_i)			0,000171	

Źródło: obliczenia własne w pakiecie LIMDEP.

Tabela 2. Oszacowania parametrów efektów grupowych oraz odchylenia od średniej

Grupa	Równanie (U)		Równanie (W)		Równanie (V)	
Grupa I	10,57	-0,03	0,40	0,01	5,44	-0,02
Grupa II	10,53	-0,07	0,42	0,03	5,40	-0,05
Grupa III	10,62	0,03	0,37	-0,02	5,48	0,03
Grupa IV	10,60	0,01	0,39	-0,00	5,46	-0,00
Grupa V	10,65	0,05	0,37	-0,02	5,49	0,04
Średnia :	10,59		0,39		5,45	

Źródło: obliczenia własne.

2. Na wynagrodzenia w krótkim okresie wpływają opóźnione o jeden kwartał stopa bezrobocia oraz wydajność pracy, natomiast w długim okresie opóźnione o dwa kwartały stopa bezrobocia i wynagrodzenia.

3. Wydajność pracy uzależniona jest w krótkim okresie od opóźnionych o jeden kwartał: wydajności pracy, wynagrodzeń i inflacji, w długim okresie zaś od opóźnionych o dwa kwartały wydajności pracy i inflacji.

Dla stopy bezrobocia najmniejszą wartość przybierają efekty grupowe w grupie II, największą zaś w grupie V. Oznacza to, że województwo lubelskie nie ma znacznego wpływu na kształtowanie się stopy bezrobocia w Polsce w badanym okresie, natomiast duży wpływ na ten wskaźnik wywierają przede wszystkim województwa zachodniopomorskie i dolnośląskie, w których stopa bezrobocia jest bardzo wysoka. Podobnie jest w przypadku równania wydajności pracy – województwo lubelskie nie ma zasadniczego wpływu na kształtowanie się poziomu wskaźnika, natomiast duży wpływ mają województwa należące do grupy V, w których wydajność pracy jest najwyższa. Dla równania wynagrodzeń najmniejszą wartość osiągają efekty grupowe w grupie V, co oznacza, iż nie mają one istotnego wpływu na kształtowanie się płac w kraju, natomiast największą w grupie II, gdzie wynagrodzenia są najniższe w kraju (podobnie jak w województwie podkarpackim).

3.2. Dynamiczny model panelowy

Ostatnim etapem analiz jest próba budowy dynamicznego modelu panelowego, tj. modelu VAR dla próby przekrojowo-czasowej. Dostępne oprogramowanie ekonometryczne (pakiet Limdep – wersja 8.0) pozwoliło estymować równania modelu za pomocą GMM. W celu analizowania szybkości dostosowań użyto małej liczby opóźnień, dlatego $k = 2$, aby nie stracić dynamicznej specyfikacji modelu. Oszacowanie parametrów nieliniowego dynamicznego modelu panelowego na próbie 496 obserwacji dało w wyniku większość parametrów nieistotnych. Efekty specyficzne dla województw zostały wyeliminowane za pomocą różnicowania zmiennych. Zaproponowano zastosowanie przyrostów sezonowych (tj. czwartych różnic dla danych kwartalnych), ostatecznie próba wyniosła 432 obserwacje w panelu.

$$\begin{aligned} \ln U_{it-4} &= \alpha_0 + \alpha_{0i} + \alpha_1 \ln(U_{-1t}) + \alpha_2 \ln(W_{-1t}) + \alpha_3 \ln(V_{-1t}) + \alpha_4 \ln(PC_{-1t}) + \\ &+ \dots + \alpha_8 \ln(PC_{-2t}) + \xi_{it} \\ \ln W_{it-4} &= \alpha_0 + \alpha_{0i} + \alpha_1 \ln(U_{-1t}) + \alpha_2 \ln(W_{-1t}) + \alpha_3 \ln(V_{-1t}) + \alpha_4 \ln(PC_{-1t}) + \\ &+ \dots + \alpha_8 \ln(PC_{-2t}) + \xi_{it} \\ \ln V_{it-4} &= \alpha_0 + \alpha_{0i} + \alpha_1 \ln(U_{-1t}) + \alpha_2 \ln(W_{-1t}) + \alpha_3 \ln(V_{-1t}) + \alpha_4 \ln(PC_{-1t}) + \\ &+ \dots + \alpha_8 \ln(PC_{-2t}) + \xi_{it}, \end{aligned} \quad (2)$$

gdzie: d – różnicowanie dla $t-4$; $i = 1, \dots, 5$ grup województw; $t = 1, \dots, 33$ okresy ($t = 1996.4, \dots, 2004.4$); U – stopa bezrobocia, PC – inflacja, W – płace przeciętne, V – wydajność pracy.

W wyniku sekwencyjnego usuwania nieistotnych parametrów otrzymano następujące wyniki:

$$\begin{aligned} \Delta \ln \hat{U}_{it} &= 0,703 \Delta \ln U_{-1it} + 0,107 \Delta \ln W_{-2it} - 1,751 \Delta \ln PC_{-1it} - 1,218 \Delta \ln PC_{-2it}, \\ \Delta \ln \hat{W}_{it} &= 0,311 + 0,340 \Delta \ln U_{-1it} - 0,501 \Delta \ln U_{-2it} + 0,376 \Delta \ln W_{-1it} + 0,287 \Delta \ln W_{-2it} - 1,550 \Delta \ln PC_{-2it}, \\ \Delta \ln \hat{V}_{it} &= 0,020 + 0,620 \Delta \ln V_{-1it} - 1,099 \Delta \ln PC_{-1it}. \end{aligned}$$

Na podstawie uzyskanych rezultatów można sformułować następujące wnioski.

1. Stopa bezrobocia jest uzależniona w krótkim okresie od opóźnionych o jeden kwartał stopy bezrobocia i inflacji, w długim okresie zaś brak wyraźnych zależności.

2. Na wynagrodzenia w krótkim okresie wpływają opóźnione o jeden kwartał stopa bezrobocia oraz wynagrodzenia, natomiast w długim okresie opóźnione o dwa kwartały: stopa bezrobocia, wynagrodzenia i inflacja.

3. Wydajność pracy uzależniona jest w krótkim okresie przede wszystkim od opóźnionych o jeden kwartał wydajności pracy i inflacji.

Wyliczone parametry posłużyły jako stałe do szacowania parametrów efektów grupowych dla poszczególnych grup województw w modelu.

Tabela 3. Oszacowania parametrów efektów grupowych oraz odchyłeń od średniej

Grupa	Równanie (U)		Równanie (W)		Równanie (V)	
Grupa I	13,72	-0,02	10,06	-0,02	8,55	-0,05
Grupa II	13,72	-0,02	10,05	-0,04	8,46	-0,14
Grupa III	13,71	-0,03	10,10	0,02	8,69	0,09
Grupa IV	13,79	0,05	10,10	0,01	8,59	-0,01
Grupa V	13,76	0,02	10,11	0,03	8,71	0,11
Średnia:	13,74		10,08		8,60	

Źródło: obliczenia własne.

Dla stopy bezrobocia najmniejszą wartość przybierają efekty grupowe w grupie III, największą zaś w grupie IV. Oznacza to, że województwa lubuskie, mazowieckie, wielkopolskie nie mają znacznego wpływu na kształtowanie się stopy bezrobocia w badanym okresie, natomiast duży wpływ na ten wskaźnik wywierają przede wszystkim województwa kujawsko-pomorskie i warmińsko-mazurskie, w którym stopa bezrobocia jest najwyższa w kraju. Dla równania wynagrodzeń najmniejszą wartość osiągają efekty grupowe w grupie II, która nie ma istotnego wpływu na kształtowanie się płac w kraju. Natomiast największą w grupie V, gdzie płace są najwyższe w kraju. Podobnie jest w przypadku równania wydajności pracy – województwo lubelskie nie ma zasadniczego wpływu na kształtowanie się poziomu wskaźnika ogólnokrajowego, natomiast duży wpływ mają województwa należące do grupy V, w których wydajność pracy jest najwyższa w kraju.

4. Podsumowanie

W opracowaniu podjęto próbę scharakteryzowania rynku pracy w Polsce, wykorzystując metody wielowymiarowej analizy porównawczej oraz modelowania wektorowo-autoregresyjnego. Zastosowanie metody aglomeracyjnej umożliwiło wyodrębnienie województw podobnych do siebie pod względem średniego poziomu badanych zmiennych: stopy bezrobocia, płac przeciętnych i wydajności pracy oraz pogrupowanie ich. Dzięki temu zredukowano wymiar danych statystycznych oraz otrzymano panel niezbilansowany, na którym przeprowadzono dalsze analizy.

Następnie, w celu opisanie rynku pracy w Polsce, oszacowano parametry modelu panelowego oraz podjęto próbę budowy dynamicznego modelu panelowego, tj. modelu VAR na próbie przestrzenno-czasowej. Zastosowanie danych panelowych do estymacji dało możliwość jednoczesnego wykorzystania informacji zarówno o dynamice, jak i o indywidualnej specyfice obserwowanych obiektów, dzięki czemu otrzymano informacje o wpływie poszczególnych grup województw na badane zmienne ekonomiczne.

Literatura

- Arellano M. (2003), *Modelling Optimal Instrumental Variables for Dynamic Panel Data Models*, Centro de Estudios Monetarios Financieros, Madrid, Working Papers No. 2003_0310.
- Charemza Wojciech W., Deadman Derek F. (1997), *Nowa ekonometria*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Hall B.H., Mairesse J., Branstetter L., Crepon B. (1998), *Does Cash Flow Cause Investment and R&D: an Exploration Using Panel Data for French, Japanese and United States Scientific Firms*, Economic Group, University of Oxford, Economic Papers No. 142.
- Hudson R.A., Owen A.L. (1996), *Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists*, Federal Reserve Board of Governors.
- Jajuga K. (1993), *Statystyczna analiza wielowymiarowa*, Wyd. Naukowe PWN, Warszawa.
- Kusideł E. (2000), *Modele wektorowo-autoregresyjne VAR metodologia i zastosowania*, [w:] *Dane panelowe i modelowanie wielowymiarowe w badaniach ekonomicznych*, t. 3, red. B. Suchecki, Absolwent, Łódź, s. 10-15.
- Malina A. (2004), *Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw*, AE, Kraków, s. 32-33.
- Sims C.A. (1980), *Macroeconomics and Reality*, „Econometrica” 48, s. 1-48.

APPLICATION OF AGGREGATION AND VECTOR-AUTOREGRESSIVE MODELING METHODS IN SPATIOTEMPORAL ANALYSIS OF LABOR MARKET IN POLAND

Summary

The results of spatiotemporal analysis of economic conditions of labor market in Poland are presented in this paper. The empirical analysis was conducted using quarterly average income, rate of unemployment, rate of inflation and work effi-

ciency data in 1997-2004. Individual characteristics of each administrative region were taken into account.

Methods of multi-scale comparison analysis were used for aggregation of statistical data. The next step was to build econometric models describing labor market in Poland using unbalanced panel model and vector-autoregressive model for spatiotemporal sample.