METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XII/2, 2011, str. 270–278

KONCEPCJA I ZASTOSOWANIE MODYFIKACJI MACIERZY WAG W PRZESTRZENNYCH BADANIACH EKONOMICZNYCH

Robert Pietrzykowski
Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: robert_pietrzykowski@sggw.pl

Streszczenie: Celem pracy jest przedstawienie zastosowania modyfikacji macierzy wag określającej interakcje przestrzenne dla ceny ziemi rolniczej w województwach. W pracy rozważono również inne zmienne o charakterze ekonomicznym, które posłużyły do określenia macierzy odległości taksonomicznych. W analizie danych wykorzystano hierarchiczną metodę analizy skupień oraz globalny współczynnik korelacji Morana. Porównano trzy rodzaje macierzy wag. W przypadku macierzy wag binarnej i unormowanej wierszowo, nie udało się stwierdzić autokorelacji przestrzennej. Natomiast stwierdzono istnienie ujemnej korelacji przestrzennej w przypadku zastosowania zmodyfikowanej macierzy wag $\mathbf{W}_m$, opartej na macierzy odległości taksonomicznych oraz macierzy sąsiedztwa.

Słowa kluczowe: korelacje przestrzenne, Nowa Ekonomia Gospodarcza (NEG), macierz wag

WSTĘP

Przestrzeń jest cechą, która powinna być uwzględniana w badaniach ekonomicznych. Jak można zauważyć, gospodarstwo jest podmiotem, który nie jest zawieszony w powietrzu, ale istnieje w konkretnym miejscu, do którego jest przypisany. W matematyce mamy możliwość poruszania się w ściśle zdefiniowanych przestrzeniach: Banacha, Hilberta, Euklidesa, kartezjańskiej, funkcyjnej i innych. W innych naukach takich jak: biologia, psychologia, politologia, socjologia, pojęcie przestrzeni odgrywa bardzo dużą rolę. Dziedziną nauki, w której przez długi okres czasu pomijano pojęcie przestrzeni była ekonomia, w której analizy przestrzenne odsyłano do geografii ekonomicznej [Kopczewska 2007]. Mogło to wynikać z wiary ekonomistów w uniwersalność praw ekonomicznych lub z braku narzędzi do analizy zjawisk ekonomicznych zlokalizowanych w przestrzeni. Uwzględnienie przestrzeni w analizach procesów ekonomicznych wymagało więc powstania nowych metod statystycznych i ekonometrycznych. Wśród twórców tej dyscypliny naukowej jaką jest ekonometria przestrzenna [Paelinck i Klaassen 1979] należy wymienić Morana, który zdefiniował współczynnik korelacji przestrzennej [Moran 1950] i Toblera [Tobler 1970], który sformułował pierwsze prawo geografii, nazwane później pierwszym prawem ekonometrii przestrzennej: „Everything is related to everything else, but near things are more related than distant things”. Choć stwierdzenie to dziś może wydawać sie trywialne, burzyło ono jednak dotychczasowe podejście do gospodarki jako niezależnego podmiotu, który nie wchodzi w interakcje przestrzenne z sąsiadami. W chwili obecnej można zauważyć dynamiczny rozwój tej dziedziny nauki na świecie, natomiast w Polsce oprócz tłumaczeń prac dotyczących tej dziedziny mamy tylko trzy publikacje książkowe polskich autorów: Zeliaś [Zeliaś 1991], Kopczewska [Kopczewska 2007] i Suchecki [Suchecki 2010]. Kluczową rolę w analizach przestrzennych odgrywa określenie wzajemnego sąsiedztwa badanych jednostek przestrzennych. Służy do tego macierz wag, w której zawarte są informacje o przestrzeni.

Celem pracy było przedstawienie zastosowania modyfikacji macierzy wag określającej interakcje przestrzenne w ekonomii.

MACIERZ WAG PRZESTRZENNYCH

Podstawowym elementem analiz przestrzennych jest określenie interakcji pomiędzy badanymi jednostkami przestrzennymi: krajami, województwami, powiatami, gminami, miastami i itd. Zgodnie z prawem Toblera jednostki przestrzenne sąsiadujące ze sobą powinny bardziej oddziaływać na siebie niż te, które znajdują się dalej. Powiązania przestrzenne należy zatem oprzeć na definicji sąsiedztwa: dwie jednostki przestrzenne uznajemy za sąsiednie jeżeli mają wspólną granicę. Powyższą definicję można zapisać w postaci macierzy binarnej w oparciu o następujący schemat działania:

$$\mathbf{W} = \begin{cases} w_{ij} = 1, & gdy\ obiekt\ i-ty\ jest\ sasiadem\ obiektu\ j-tego \\ w_{ij} = 0, & gdy\ obiekt\ i-ty\ nie\ jest\ sasiadem\ obiektu\ j-tego \\ w_{ij} = 0, & gdy\ i = j \end{cases} \qquad (1)$$

Tak skonstruowaną macierz, która określa sąsiedztwo poszczególnych jednostek przestrzennych określa się jako macierz wag (**W** - formuła 1), natomiast w_{ij} są elementami macierzy sąsiedztwa. Nie zawsze kryterium wspólnej granicy może dobrze odzwierciedlać wzajemne powiązania między badanymi obiektami, dlatego często, wybiera się inne kryteria w celu uzyskania macierzy sąsiedztwa np. wybierając sąsiadów w odległość d km. Innym sposobem określenia elementów macierzy wag **W** jest przyjęcie interakcji przestrzennych poprzez wprowadzenie odległości społecznych [Doreian 1980] lub ekonomicznych [Conley 1999], które oparte są na wzajemnych stosunkach handlowych, przepływie kapitału oraz migracjach pomiędzy badanymi jednostkami przestrzennymi (obiektami). Inną propozycją macierzy wag jest macierz uwzględniająca długość granicy sąsiadujących ze sobą jednostek przestrzennych (obiektów) [Dacey'a 1968].

Jak można zauważyć określenie macierzy wag w analizach przestrzennych jest bardzo ważne. Dyskusję na temat określania macierzy wag można znaleźć u Clifa i Orda [Clif i Ord 1981], Anselina [Anselin 1988], Uptona i Fingeltona [Upton i Fingelton 1985]. Ponieważ jak już wspomniano wcześniej, macierz wag określa się *a priori*, a jakość dalszych analiz zależy od jej specyfikacji. Dobór odpowiedniej macierzy wag stanowi poważny problem metodologiczny. Najczęściej stosowana macierz wag to macierz oparta na kryterium wspólnej granicy, pierwszego rzędu oraz tak przekształcona, aby suma elementów każdego wiersza wynosiła 1.

W pracy zaproponowano modyfikację macierzy wag **W**, która uwzględniała by interakcje związane z położeniem badanej jednostki przestrzennej, ale również miała wagi związane z wybranymi cechami ekonomicznymi. W tym celu w pierwszym etapie badań na bazie *m* cech o charakterze ekonomicznym, określono macierz odległości taksonomicznych, które obliczono dla badanych jednostek przestrzennych (województw) zgodnie z formułą:

$$d_{ik} = \left[\sum_{j=1}^{m} \left| x_{ij} - x_{kj} \right|^{p} \right]^{\frac{1}{p}} \tag{2}$$

gdzie: p jest liczbą określającą rodzaj metryki,
m jest liczbą cech,
x_{ij}, x_{kj} określają realizację j – tej cechy w obiektach i – tym oraz k – tym.

Uzyskane odległości d_{ik} zestawiono w macierzy odległości **O**. Macierz odległości podobnie jak macierz wag na diagonali ma zera i jest macierzą symetryczną.

$$\mathbf{W}_n = \mathbf{W}_s + \mathbf{O} \tag{3}$$

gdzie:

$\mathbf{W}_n$ – zmodyfikowana macierz wag
$\mathbf{W}_s$ – macierz wag pierwszego rzędu
O – macierz odległości taksonomicznych

Macierz wag **W** uzyskano zgodnie z formułą 1, a następnie unormowano wierszowo. Uzyskano w ten sposób macierz sąsiedztwa pierwszego rzędu, czyli uwzględniającą tylko najbliższych sąsiadów (rysunek 1). Tak uzyskana macierz wag nie jest macierzą symetryczną, dlatego po dodaniu wartości tej macierzy do macierzy odległości taksonomicznych $\mathbf{W}_n$ poddano transformacji w celu zapewniania prawidłowości wnioskowania na podstawie aproksymowanych momentów z próby [Anselin 2001] zgodnie z formułą:

$$\mathbf{W}_m = \frac{(\mathbf{W}_n + \mathbf{W}_n^T)}{2} \tag{4}$$

gdzie $\mathbf{W}_n$ jest macierzą wag uzyskaną według formuły 3

Rysunek. 1. Przykład powiązań według kryterium wspólnej granicy uwzględniający tylko sąsiedztwo pierwszego rzędu

Źródło: opracowanie własne

Nowa macierz wag $\mathbf{W}_m$ jest zatem macierzą, która ma na diagonali zera i jest macierzą symetryczną. W swojej strukturze uwzględnia odległości związane z badanymi cechami, które pośrednio mogą wpływać na interesującą nas cechę oraz wpływ sąsiada. W dalszej części pracy obliczono globalny współczynnik korelacji Morana dla trzech macierzy wag $\mathbf{W}_s$, $\mathbf{W}_b$ i $\mathbf{W}_m$. Macierz $\mathbf{W}_b$ jest macierzą binarną, określającą wzajemne sąsiedztwo obiektów, uzyskaną zgodnie z formułą 1. Macierz $\mathbf{W}_s$ jest macierzą, którą uzyskano po przekształceniu macierzy $\mathbf{W}_b$, tak aby suma elementów każdego wiersza dawała 1. Natomiast macierz $\mathbf{W}_m$ jest zmodyfikowaną macierzą wag, która oprócz informacji o wzajemnym sąsiedztwie obiektów (sąsiedztwo pierwszego rzędu – formuła 1), zawierała również powiązania obiektów związane z wybranymi cechami ekonomicznymi, które uzyskano poprzez macierz odległości taksonomicznych (zgodnie z formułami 2, 3 i 4).

AUTOKORELACJA PRZESTRZENNA

Globalny współczynnik korelacji Morana, służy do badania zależności przestrzennych dla danej zmiennej w obrębie całego badanego obszaru. Można go również traktować jako miernik grupowania badanych jednostek w przestrzeni. Konstrukcja tego miernika wykorzystuje koncepcję współczynnika korelacji Pearsona [Sang-Il 2001] oraz statystykę gamma, która łączy dwa rodzaje informacji dotyczące podobieństw obserwacji w przestrzeni. Informacje te zawarte są w macierzy wag **W** oraz w macierzy korelacji **A**, które dotyczą zależności badanej zmiennej w poszczególnych lokalizacjach. Ogólną postać statystyki gamma możemy zapisać:

$$\Gamma = \mathbf{WA} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} a_{ij} \tag{5}$$

gdzie: w_{ij} – elementy macierzy wag **W**
a_{ij} – elementy macierzy **A**.

Statystykę Morana można zatem zapisać jako:

$$\Gamma = \sum_{i}\sum_{j} w_{ij}(z_i \cdot z_j) \tag{6}$$

gdzie z_i i z_j są standaryzowanymi wartościami zmiennej losowej x.

Biorąc pod uwagę dowód Clifa i Orda [1973] globalna statystyka korelacji przestrzennej Morana ma rozkład asymptotycznie normalny, a jej zapis macierzowy przedstawiono w formule 7.

$$I_g = \frac{n}{S_0}\frac{\mathbf{z}'\mathbf{W}\mathbf{z}}{\mathbf{z}'\mathbf{z}} \tag{7}$$

$$Z(I_g) = \frac{I_g - E(I_g)}{\operatorname{var}(I_g)^{\frac{1}{2}}} \sim \mathrm{N}(0,1) \tag{8}$$

gdzie: z jest wektorem o elementach $z_i = x_i - \bar{x}$
S_0 jest sumą wszystkich elementów macierzy wag

Wykorzystując współczynnik korelacji Morana można weryfikować następującą hipotezę: H_0: obserwowane wartości badanej zmiennej są rozmieszczone w sposób losowy pomiędzy poszczególnymi jednostkami przestrzennymi (brak autokorelacji przestrzennej), wobec hipotezy alternatywnej: H_1: obserwowane wartości badanej zmiennej nie są rozmieszczone w sposób losowy pomiędzy poszczególnymi jednostkami przestrzennymi (występuje autokorelacja przestrzenna). Weryfikację hipotezy zerowej przeprowadza się w oparciu o statystykę $Z(I_g)$ (formuła 8). Przyjmuje się, że w przypadku braku autokorelacji przestrzennej statystyka Morana Ig ma tendencje do przyjmowania wartości $I_g \approx E(I_g)$, $Z(I_g) \approx 0$. Natomiast w przeciwnym razie występuje: autokorelacja dodatnia i wtedy $I_g > E(I_g)$, $Z(I_g) > 0$ lub autokorelacja ujemna, gdy wartości $I_g < E(I_g)$, $Z(I_g) < 0$.

Autokorelacje przestrzenne można również przedstawić na wykresie rozrzutu Morana, który dotyczy wartości współczynnika globalnego Morana (Ig). Wykres ten jest wykorzystywany do wizualizacji związków przestrzennych i określenia kierunku autokorelacji przestrzennej. Wykres podzielony jest na cztery części względem wartości zerowych. Natomiast na osiach naniesione są standaryzowane wartości badanej zmiennej oraz opóźnienie badanej zmiennej. Do otrzymanego zbioru punktów dopasowujemy prostą, a współczynnik kierunkowy prostej jest równoważny wartości współczynnika korelacji Morana [Kopczewska 2007, Suchecki 2010].

Wadą wszystkich statystyk globalnych autokorelacji przestrzennej jest to, że ich wartość zależy od skali podziału badanego regionu. W przypadku silnej koncentracji, wyniki przy różnych podziałach czyli różnych macierzach wag **W,** mogą różnić się diametralnie.

WYNIKI

W dalszej części pracy zbadano zależność przestrzenną dla ceny ziemi rolniczej. Rozważano następujące zmienne: ceny gruntów w obrocie prywatnym, jednolite dopłaty bezpośrednie, uzupełniające dopłaty bezpośrednie, dopłaty do ONW, nawożenie mineralne w kg/ha, powierzchnia użytków rolnych według klas bonitacyjnych w hektarach (klasy V, VI). Wszystkie zmienne obserwowano na poziomie województw (NUT 2). Dane pochodziły z GUS i ANR z roku 2009. Wyniki analiz przestrzennych dla różnych macierzy wag przedstawiono w tabeli 1.

Tabela 1. Statystyka Morana Ig dla różnych macierzy wag

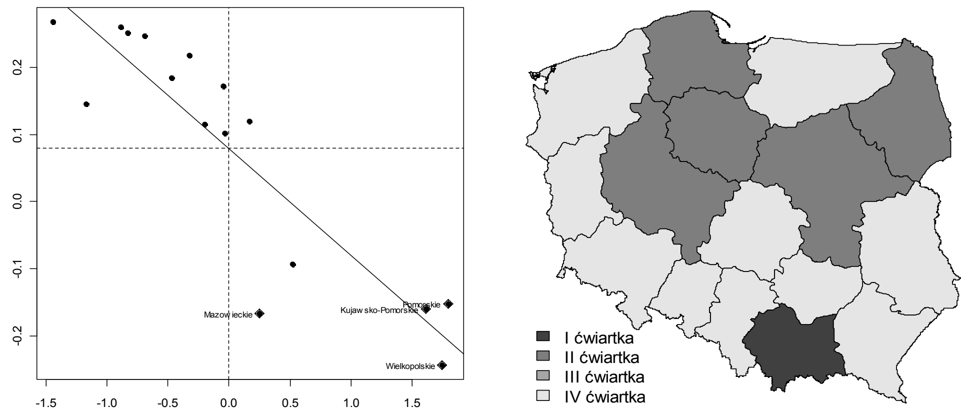
Macierz wag	Ig	E(Ig)	var(Ig)	Z(Ig)	Poziom krytyczny
$\mathbf{W}_m$	-0,2209	-0,0667	0,0014	-4,1239	3,724e-05
$\mathbf{W}_b$	-0,1118	-0,0667	0,0195	-0,3235	0,7463
$\mathbf{W}_s$	-0,0922	-0,0667	0,0219	-0,1726	0,8630

Źródło: opracowanie własne

Przy zastosowaniu trzech różnych macierzy wag (binarnej - $\mathbf{W}_b$, unormowanej wierszowo - $\mathbf{W}_s$ i uwzględniającej odległości taksonomiczne oraz sąsiedztwo obiektów - $\mathbf{W}_m$) tylko dla macierzy $\mathbf{W}_m$ współczynnik korelacji był istotny (poziom krytyczny: 3,724e-05 oraz wartość Ig i Z(Ig) są mniejsze od zera). Oznacza to, że dla badanej ceny ziemi stwierdzono korelację przestrzenną, czyli odrzucono hipotezę, że obserwowane wartości cen ziemi rolniczych są rozmieszczone w sposób losowy pomiędzy województwami. A zatem możemy stwierdzić, że obserwowane wartości cen ziemi rolniczej nie są rozmieszczone w sposób losowy pomiędzy województwami i tworzą reżimy przestrzenne ze względu na badaną cechę. Należy również pamiętać, że macierz $\mathbf{W}_m$ zawierała informacje o interakcjach pomiędzy województwami ze względu na wybrane cechy ekonomiczne jakimi były: jednolite dopłaty bezpośrednie, uzupełniające dopłaty bezpośrednie, dopłaty do ONW, nawożenie mineralne w kg/ha, powierzchnia użytków rolnych według klas bonitacyjnych w hektarach dla klas V i VI. Na wykresie Morana przedstawiono rozmieszczenie województw w czterech ćwiartkach (rysunek 2, lewy).

Na przedstawionym wykresie (rysunek 2, lewy) wyraźnie widać występującą ujemną autokorelację przestrzenną. Województwa zgrupowane są w drugiej i czwartej ćwiartce. W czwartej ćwiartce znajduje się dziesięć województw i są to województwa: zachodniopomorskie, lubuskie, opolskie, śląskie, dolnośląskie, świętokrzyskie, łódzkie, podkarpackie, lubelskie i warmińsko-mazurskie.

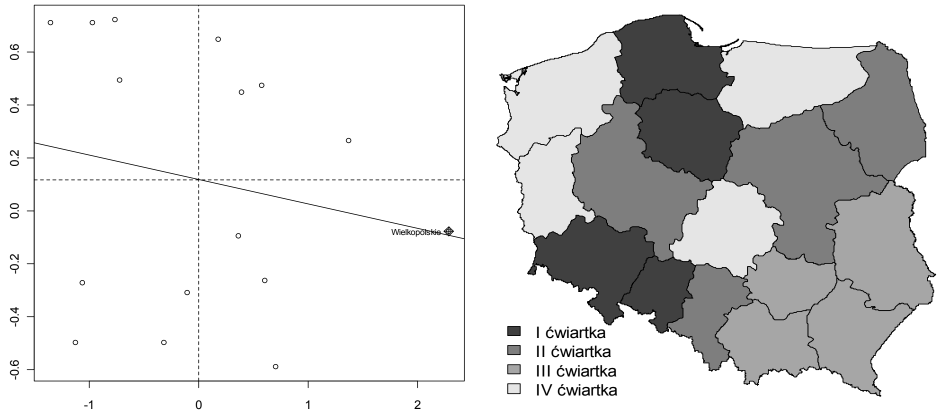
Rysunek 2. Wykres rozrzutu Morana (lewy) i jego przestrzenna wizualizacja (prawy)



Źródło: opracowanie własne

W pierwszej ćwiartce znajduje się jedno województwo: małopolskie i to województwo można traktować jako punkt odstający. Pozostałe pięć województw znajduje się w drugiej ćwiartce i są to: pomorskie, kujawsko-pomorskie, wielkopolskie, mazowieckie i podlaskie. Na rysunku 2 (prawy) przedstawiono na mapie Polski rozmieszczenie województw według ćwiartek z wykresu Morana. Widać, że województwa nie są rozmieszczone w sposób losowy ze względu na badaną cechę. Na wykresie trzecim przedstawiono przypadek braku autokorelacji w przypadku zastosowania macierz wag unormowanej wierszami W_s

Rysunek 3. Wykres rozrzutu Morana (lewy) i jego przestrzenna wizualizacja (prawy) dla macierzy wag W_s



Źródło: opracowanie własne

Widać, że punkty na wykresie rozrzutu Morana są rozmieszczone dość równomiernie pomiędzy cztery ćwiartki, czyli potwierdza to wcześniejsze badania, w których wykazano brak autokorelacji przestrzennej (poziom krytyczny: 0,8630). Również na mapie wojewódzkiej Polski widać, że rozmieszczenie cen ziemi rolniczej w poszczególnych województwach ma charakter „szachownicy", czyli nie można stwierdzić zależności przestrzennych.

PODSUMOWANIE

Wykorzystanie modyfikacji macierzy wag pozwoliło na dokładniejsze zbadanie zależności przestrzennych. Macierz wag $\mathbf{W}_m$ uwzględnia nie tylko wzajemne interakcje związane z sąsiedztwem jednostek przestrzennych, ale również interakcje które wynikają z innych cech bezpośrednio nie związanych z daną cechą lub jednostką przestrzenną. Wykorzystując macierze wag binarną i unormowaną wierszami, które zawierały tylko informację o sąsiadach pierwszego rzędu nie udało się stwierdzić autokorelacji przestrzennej. W przypadku zastosowania macierzy wag $\mathbf{W}_m$, stwierdzono ujemną autokorelację przestrzenną oznaczająca rywalizację przestrzenną, czyli wysokie wartości cen ziemi rolniczej odpowiadają niskim wartościom cen ziemi rolniczej w sąsiednich lokalizacjach (województwach) i odwrotnie. Tak uzyskaną macierz wag ($\mathbf{W}_m$) można wykorzystać w dalszych badaniach dla modeli regresji przestrzennej. W pracy przedstawiono tylko analizę dotyczącą globalnego współczynnika korelacji Morana, pomijając analizę lokalnych współczynników korelacji oraz analizę regresji przestrzennej.

BIBLIOGRAFIA

Anselin L. (1988) Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity, Geographical Analysis, 20, s. 1 - 17

Anselin L. (2001) Spatial econometrics, Oxford, Basil Blackwell

Cliff A. D., Ord K. (1981): Spatial Process: Models and Applications, Pion, London

Conley T. G. (1999): GMM estimation with cross selection dependence, Journal of Econometrics, 92(1), s. 1 – 45

Dacey M. (1968) A review of measures of contiguity for two and k-color maps. In spatial analysis: A Reader in Statistical Geography, B. Berry and D. Marble (eds), Englewood Cliffs, N.J., Prentics-Hall, s. 479 – 495

Doreian P. (1980) Lineral models with spatial distributed data. Spatial disturbances or spatial effects, Sociological Methods and Research, 9, 29 – 60

Kopczewska K. (2007) Ekonometria i statystyka przestrzenna, CEDEWU.PL, Warszawa

Moran, P.A.P. (1950) Notes on Continuous Stochastic Phenomena, Biometrika, 37, 17–33.

Paelinck J, Klaassen L (1979) Spatial econometrics. Saxon House, Farnborough

Sang-Il L. (2001) Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I, Journal of Geographical System vol 3, s. 369 - 385

Suchecki B. (2010) Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych, C.H. Beck, Warszawa

Tobler W. (1970) A computer movie simulating urban growth in the Detroit region. Economic Geography, 46(2): 234-240

Upton G., Fingleton B. (1985) Spatial Data Analysis by Example, Wiley, New York

Zeliaś A. (1991): Ekonometria przestrzenna, PWN, Warszawa

CONCEPT AND APPLICATION MATRIX WEIGHT MODIFICATIONS IN SPATIAL ECONOMIC RESEARCH

Abstract: The aim of this study was to present the application of modified weight matrix defining the spatial interactions for the price of agricultural land in the provinces. The study also considered other variables of an economic nature, which were used to determine the distance matrix. The data analysis used a hierarchical cluster analysis method and the global correlation Moran's coefficient. Compared the three types of matrix weight. In the case of binary weights matrix and rows normalized, failed to determine spatial autocorrelation. However there is a negative correlation was found in the case of spatial use a modified weight matrix W_m.

Key words: spatial correlations, The New Economic Geography (NEG), weight matrix