RUCH PRAWNICZY, EKONOMICZNY I SOCJOLOGICZNY
Rok XLV — zeszyt 3 — 1983

EDWARD NOWAK

REGRESYJNE BADANIA EFEKTYWNOŚCI PRODUKCJI W WARUNKACH NIEJEDNORODNOŚCI ZBIORU OBIEKTÓW

I. UWAGI WSTĘPNE

W badaniach zależności między nakładami na produkcję a jej efektami często wykorzystywana jest analiza regresji. Jednym z podstawowych złożeń, które muszą być spełnione przy stosowaniu metod analizy regresji, decydującym o poprawności interpretacji i wykorzystania wyników, jest jednorodność zbioru obiektów. Warunek ten nabiera szczególnego znaczenia w badaniach przekrojowych, gdy obiektami badania nie są jednostki czasu, lecz przykładowo przedsiębiorstwa, zakłady, zatrudnieni, jednostki terytorialne, itp., które najczęściej tworzą zbiór niejednorodny ze względu na opisujące je zmienne, tj. nakłady na produkcję i efekty produkcji.

Jeśli istnieje przypuszczenie, że obiekty badania tworzą niejednorodny zbiór ze względu na opisujące je zmienne, wówczas należy przeprowadzić jego podział na grupy typologiczne obejmujące obiekty podobne ze względu na nakłady na produkcję i efekty produkcji, a modele regresyjne budować dla poszczególnych grup[1]. W różnych grupach typologicznych obejmujących obiekty jednorodne mogą bowiem istnieć różne powiązania między wyróżnionymi zjawiskami: nakłady mogą mieć różny wpływ na kształtowanie się poziomu produkcji.

W warunkach niejednorodności zbioru obiektów model regresji budowany na podstawie informacji dotyczących wszystkich obiektów posiada jedynie wartości poznawcze jako opis całego zbioru badanych obiektów. Nie posiada natomiast wartości predyktywnych i wartości poznawczych jako opis relacji nakłady-efekty w ramach wydzielonych grup.

Podział zbioru wyróżnionych obiektów na grupy obiektów jednorodnych jest czynnością wstępną dla właściwych analiz zależności między poziomem produkcji a wysokością nakładów. Badania regresyjne efektywności produkcji dotyczą powiązań między dwoma odrębnymi, jakoś-

[1] Problem ten jest szeroko omawiany w pracy B. Rozina, *Teoria rozpoznawania obrazów w badaniach ekonomicznych,* Warszawa 1976.

ciowo odmiennymi zjawiskami dotyczącymi różnych stron procesu wytwórczego. Jedno zjawisko to wyniki produkcyjne opisywane przez zmienną objaśnianą, drugie zjawisko to nakłady i zasoby czynników produkcji reprezentowane przez zmienne objaśniające. Fakty te powinny być uwzględnione przy przeprowadzaniu dyskryminacji zbioru obiektów badania; procedura dyskryminacji powinna uwzględniać istnienie dwóch różnych kryteriów podziału: zmiennej objaśnianej z jednej strony i zmiennych objaśniających z drugiej strony. Dla celów niniejszego opracowania przyjmijmy, że podział ten został dokonany[2].

Klasyczne podejście do grupowych badań regresyjnych sprowadza się do wyznaczenia równań regresji dla poszczególnych grup typologicznych na podstawie danych statystycznych dotyczących tych grup. Oznacza to, że nie uwzględnia się istnienia pewnych wspólnych dla wszystkich obiektów badanej zbiorowości związków łączących efekty produkcyjne z nakładami na produkcję, które z reguły, nawet w przypadku zbiorowości niejednorodnych, istnieją. Powstaje przeto potrzeba uwzględnienia tego faktu przy prowadzeniu grupowych badań regresyjnych, co będzie dalej zaproponowane w niniejszej pracy.

Oznaczmy przez

$$\Omega=\{P_1, P_2, \ldots, P_n\} \tag{1}$$

zbiór badanych obiektów. Niech dalej Y oznacza produkcję, natomiast

$$X=\{X_1, X_2, \ldots, X_K\} \tag{2}$$

zbiór zmiennych opisujących nakłady i zasoby czynników produkcji.

Obserwacje dotyczące powyższych wielkości można przedstawić w postaci:

— wektora obserwacji zmiennej objaśnianej:

$$\mathbf{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \tag{3}$$

gdzie: $y_t (t = 1, 2, \ldots, n)$ oznacza poziom produkcji w obiekcie P_t,

—- macierzy obserwacji zmiennych objaśniających:

$$\mathbf{X}=\begin{bmatrix} x_{11} & x_{21} & \ldots & x_{K1} \\ x_{12} & x_{22} & \ldots & x_{K2} \\ \cdot & \cdot & \cdot & \cdot \\ x_{1n} & x_{2n} & \ldots & x_{Kn} \end{bmatrix}, \tag{4}$$

[2] O podziale zbioru obiektów dla badań regresyjnych szeroko traktuje praca zbiorowa pod kierunkiem S. Bartosiewicz, *Zastosowanie technik wielowymiarowej analizy porównawczej w dynamicznych i przekrojowych badaniach ekonomicznych^ Etap I. Problem jednorodności zbiorów obiektów i zbiorów cech,* Akademia Ekonomiczna, Wrocław 1901 (Problem resortowy R. III. 9).

gdzie: x_{kt} $(k=1, 2, \ldots, K;\ t=1, 2, \ldots, n)$ oznacza nakłady k-tego czynnika produkcji dla obiektu P_t.

Załóżmy, że w wyniku podziału zbioru obiektów Ω na grupy obiektów podobnych ze względu na wyniki produkcyjne oraz nakłady i zasoby czynników produkcji otrzymano G grup typologicznych:

$$\Omega=\{C_1, C_2, \ldots, C_G\}. \tag{5}$$

Niech dalej i oznacza numer grupy typologicznej: $i= 1, 2, \ldots, G$.

II. DWUSTOPNIOWE REGRESJE GRUPOWE

Poczynione wcześniej uwagi wskazują, że można sformułować dwa zasadniczo odmienne podejścia do regresyjnego modelowania produkcji względem nakładów. Podejście pierwsze przyjmuje, że za model zmiennej Y przyjmuje się globalną regresję

$$\bar{Y}=a_1 X_1+a_2 X_2+\ldots+a_K X_K, \tag{6}$$

wyznaczoną na podstawie danych statystycznych dotyczących wszystkich obiektów zbioru Ω. Takie rozwiązanie jest uzasadnione przypuszczeniem, że wszystkie obiekty dobrze opisuje się jedną funkcją regresji.

W sytuacji, gdy obiekty badania tworzą niejednorodny zbiór ze względu na opisujące je zmienne (efekty i nakłady) podejście to nie jest uzasadnione. Z tego powodu stosuje się rozwiązanie alternatywne sprowadzające się do wyznaczenia regresji dla poszczególnych grup typologicznych C_i na podstawie danych statystycznych dotyczących obiektów tworzących te grupy. Oznaczmy liczebności poszczególnych grup typologicznych przez n_i $(i=1, 2, \ldots, G)$. Tak więc dla grupy C_i $(i=1, 2, \ldots, G)$ na podstawie informacji dotyczących zmiennej objaśnianej

$$\mathbf{Y}_i=\begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{n_i i} \end{bmatrix}, \tag{7}$$

i zmiennych objaśniających

$$\mathbf{X}_i=\begin{bmatrix} x_{11i} & x_{21i} & \cdots & x_{K1i} \\ x_{12i} & x_{22i} & \cdots & x_{K2i} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1n_i i} & x_{2n_i i} & \cdots & x_{Kn_i i} \end{bmatrix} \tag{8}$$

wyznacza się regresję

$$\hat{Y}_i=a_{1i} X_{1i}+a_{2i} X_{2i}+\ldots+a_{Ki} X_{Ki}, \tag{9}$$

$$i=1, 2, \ldots, G.$$

W ten sposób wyznaczone regresje grupowe opisują relacje między produkcją a nakładami właściwe dla poszczególnych grup typologicznych. W tym przypadku za model zmiennej Y przyjmuje się funkcję regresji składającą się z grupowych funkcji regresji wyznaczonych oddzielnie dla poszczególnych grup.

Regresja globalna (6) opisuje zmienność zmiennej objaśnianej wynikającą z własności charakterystycznych dla całego zbioru obiektów. Z kolei regresje grupowe (9) wyjaśniają zmienność zmiennej objaśnianej w poszczególnych grupach typologicznych wynikającą ze specyficznych własności tych grup.

Przedstawione postępowanie sprowadzające się do podziału zbioru obiektów na grupy typologiczne i wyznaczanie funkcji regresji odrębnych dla każdej grupy nie zawsze jest wskazane i uzasadnione, mimo iż często jest zalecane i stosowane w badaniach praktycznych. Postępowanie takie zakłada bowiem, że istnieją specyficzne prawidłowości właściwe jedynie dla poszczególnych grup typologicznych, że jedynie w grupie obiektów podobnych można zbadać związek zmian zmiennej objaśnianej ze zmianami zmiennych objaśniających. Z reguły jednakże jest tak, że dla całego badanego zbioru występuje ogólna zależność zmiennej objaśnianej od zmiennych objaśniających, a ponadto istnieją specyficzne własności poszczególnych grup obiektów. W takich przypadkach tradycyjny sposób prowadzenia grupowych badań regresyjnych nie zawsze prowadzi do zadowalających rezultatów. Konieczne jest przeto zastosowanie rozwiązania pośredniego. Niekiedy wskazane jest w taki sposób budować grupowe funkcje regresji, by uwzględniały zarówno właściwości całego zbioru obiektów, jak i specyfikę grup, by opisywały zmiany zmiennej objaśnianej wynikające zarówno z globalnych własności całego zbioru obiektów, jak i spowodowane specyficznymi własnościami grup obiektów podobnych. Przedstawiona niżej propozycja jest próbą rozwiązania tego zagadnienia.

Punktem wyjścia proponowanego rozwiązania jest regresja ogólna (6) szacowana dla całej zbiorowości obiektów, z której wyznaczane są teoretyczne wartości zmiennej objaśnianej

$$\bar{y}_t = a_1 x_{1t} + a_2 x_{2t} + \ldots + a_K x_{Kt}, \quad t = 1, 2, \ldots, n. \tag{10}$$

Oznaczmy przez

$$\mathbf{Y}_i = \begin{bmatrix} \bar{y}_{1i} \\ \bar{y}_{2i} \\ \vdots \\ \bar{y}_{n_i i} \end{bmatrix}, \quad i = 1, 2, \ldots, G, \tag{11}$$

wektor teoretycznych wartości zmiennej objaśnianej Y dla i-tej grupy wyznaczonych z ogólnej funkcji regresji, a więc na podstawie wartości zmiennych objaśniających z tej grupy:

$$\bar{y}_{ti} = a_1 x_{1ti} + a_2 x_{2ti} + \ldots + a_K x_{Kti}, \qquad (12)$$

$$i = 1, 2, \ldots, G; \quad t = 1, 2, \ldots, n_i.$$

W następnym etapie dla każdej grupy typologicznej szacowana jest grupowa funkcja regresji. W tym przypadku rozszerza się zbiór danych (7) i (8) dotyczących obiektów danej grupy o teoretyczne wartości zmiennej objaśnianej (11) tej samej grupy wyznaczone z globalnej funkcji regresji i wartości zmiennych objaśniających (8). Liczba obserwacji wykorzystanych do szacowania regresji grupowych zwiększa się więc dwukrotnie i wynosi teraz $2n_i$.

Oszacowane w ten sposób grupowe funkcje regresji można przedstawić następująco:

$$\begin{pmatrix} \hat{Y}_i \\ \bar{Y}_i \end{pmatrix} = \hat{a}_{1i} \begin{pmatrix} X_{1i} \\ X_{1i} \end{pmatrix} + \hat{a}_{2i} \begin{pmatrix} X_{2i} \\ X_{2i} \end{pmatrix} + \ldots + \hat{a}_{Ki} \begin{pmatrix} X_{Ki} \\ X_{Ki} \end{pmatrix}, \qquad (13)$$

$$i = 1, 2, \ldots, G.$$

Równania powyższe można nazwać d w u s t o p n i o w y m i r e g r e s j a m i g r u p o w y m i. Regresje te wyjaśniają zmiany zmiennej objaśnianej Y w poszczególnych grupach typologicznych wynikające zarówno z własności całego zbioru badanych obiektów, jak i z własności specyficznych tych grup. W tym sensie dwustopniowe regresje grupowe stanowią kompromis między regresją ogólną dotyczącą całego zbioru badanych obiektów a regresjami grupowymi wyznaczanymi w tradycyjny sposób.

W literaturze można znaleźć odmienne rozwiązania dotyczące uwzględniania informacji o całym zbiorze obiektów przy prowadzeniu badań regresyjnych w przekroju grup typologicznych. Niekiedy wychodzi się z założenia, że mimo podziału zbioru obiektów na grupy typologiczne grupy te w dalszym ciągu są między sobą w mniejszym lub większym stopniu powiązane. Dlatego niekiedy przy wyznaczaniu regresji grupowych wykorzystuje się także informacje z innych grup brane z odpowiednią wagą, w zależności od stopnia podobieństwa grup. Uzyskuje się przez to zwiększenie liczebności zbioru danych, co umożliwia szacowanie regresji nawet dla mało licznych grup. Rozwiązania te można także odnieść do zagadnienia poruszonego w niniejszej pracy.

Jedno z tych rozwiązań[3] to wyznaczanie funkcji regresji dla poszczególnych grup typologicznych na podstawie danych pochodzących ze

[3] Por. A. W. Bekker, M. A. Jagolnicer, A. A. Kołokołow, B. A. Gładkich, *Raspoznawanije obrazow pri postrojenii ekonomiko-statisticzeskich modieliej,* Nowosybirsk 1975.

wszystkich grup, przy czym obserwacje te przekształcane są w ten sposób, że

$$y^{*}_{tj} = w_{ij} \cdot y_{tj}, \tag{14}$$

$$x^{*}_{ktj} = w_{ij} \cdot x_{ktj}, \tag{15}$$

gdzie: w_{ij} — unormowany na [0,1] wskaźnik podobieństwa obiektów i-tej grupy z pozostałymi grupami obiektów; y^{*}_{tj} — ważone obserwacje zmiennej objaśnianej z grupy *j*-tej; x^{*}_{ktj} — ważone obserwacje zmiennych objaśniających z grupy *j*-tej. Obserwacje grupy, dla której szacuje się równanie regresji brane są z wagą równą jedności.

Wadą takiego rozwiązania jest, iż niejako wymusza ono przechodzenie grupowych funkcji regresji przez początek lub blisko początku układu współrzędnych, a to z kolei powoduje znaczne różnice w wartościach ocen parametrów ważonych regresji grupowych w stosunku do wartości ocen parametrów klasycznych regresji grupowych i regresji globalnej.

Inne postępowanie[4] polega na zwielokrotnianiu obserwacji poszczególnych grup w zależności od stopnia podobieństwa do wyróżnionej grupy, dla której szacuje się regresję. Wskaźnik wyznaczający liczbę powtórzeń obserwacji danej grupy ma postać:

$$N_{ij} = 10^{m} \cdot w_{ij}, \tag{16}$$

gdzie: m — zadana z góry liczba naturalna: $m = 1, 2, \ldots$, w_{ij} — wskaźnik podobieństwa grup C_i i C_j, N_{ij} — liczba zwielokrotnienia liczebności grupy o numerze *j* przy szacowaniu regresji dla grupy *i*-tej.

Obserwacje grupy, dla której szacuje się równanie regresji zwielokratniane są 10^{m} razy.

Wyznaczone w *ten* sposób tzw. regresje autonomiczne są pośrednimi między regresjami grupowymi szacowanymi w tradycyjny sposób a ogólną regresją wyznaczoną dla całej zbiorowości obiektów. Wadą takiego postępowania są natomiast trudności związane z ustalaniem parametru m we wzorze (16), która to wielkość jest zadawana z góry przez badacza, a od której bardzo mocno zależy liczba powtórzeń obserwacji grupy. Dla mniejszego *m* regresja autonomiczna zbliża się bardziej do regresji ogólnej, natomiast dla m większego bardziej od tej regresji oddala się. Z drugiej strony, większe wartości m powodują zbliżanie się regresji autonomicznych do klasycznych regresji grupowych.

[4] Por. praca zbiorowa pod kierunkiem S. Bartosiewicz, *Zastosowanie technik wielowymiarowej analizy porównawczej w dynamicznych i przekrojowych badaniach ekonometrycznych. Etap II. Problem separowalności zbiorów obiektów i zbiorów cech. Autonomiczne funkcje quasi-regresji,* Akademia Ekonomiczna, Wrocław 1982 (Problem resortowy R. III. 9).

III. EGZEMPLIFIKACJA PREZENTOWANEJ KONCEPCJI

Przedstawiona propozycja wyznaczania dwustopniowych regresji grupowych zostanie zilustrowana przykładem empirycznym dotyczącym 30 zakładów produkcyjnych. Ze względu na możliwość interpretacji graficznej ograniczono się do regresji z jedną zmienną objaśniającą. Rolę zmiennej objaśnianej pełni produkcja wyrażana w sztukach, natomiast zmienną objaśniającą jest zatrudnienie w osobach. Zbiór badanych zakładów został podzielony na trzy grupy typologiczne obejmujące zakłady podobne ze względu na poziom produkcji i poziom zatrudnienia. Wartości zmiennej objaśnianej i zmiennej objaśniającej w poszczególnych zakładach produkcyjnych zawarte są w tabeli 1.

Tabela 1

Produkcja i zatrudnienie w poszczególnych zakładach

Numer zakładu	Grupa I		Grupa II		Grupa III	
	Y	X	Y	X	Y	X
1	480	107	518	147	581	192
2	490	110	530	156	592	191
3	464	119	517	162	595	200
4	475	125	532	167	617	201
5	447	128	499	172	620	210
6	449	136	521	179	600	218
7	467	137	539	176	624	220
8	434	148	532	189	640	217
9	438	156	543	188	655	228
10	422	162	542	205	670	227

Źródło: Dane umowne.

Na podstawie danych statystycznych dotyczących wszystkich 30 zakładów produkcyjnych oszacowana została globalna funkcja regresji produkcji względem zatrudnienia. W wyniku tego otrzymano następujące równanie regresji:

$$\hat{Y} = 1{,}694X + 242{,}3,$$

gdzie: Y — produkcja w tys. sztuk, X — zatrudnienie w osobach.

Na podstawie danych zawartych w tabeli 1 oszacowano także klasyczne regresje dla poszczególnych grup typologicznych. Równania te przedstawiają się następująco:

— równanie grupy I:

$$\hat{Y}_1 = -1{,}07X + 598{,}8,$$

— równanie grupy II:

$$\hat{Y}_2 = 0{,}409X + 456{,}1,$$

— równanie grupy III:

$$\hat{Y}_3 = 1{,}783X + 244{,}1$$

Kształtowanie się regresji globalnej i regresji grupowych przedstawione jest na rycinie.

Dla wyznaczenia dwustopniowych regresji grupowych policzone zostały teoretyczne wartości zmiennej objaśnianej z globalnej funkcji regresji. Wartości te zawarte są w tabeli 2.

Tabela 2

Teoretyczne wartości produkcji

Numer zakładu	Grupa I	Grupa II	Grupa III
1	424	491	568
2	429	507	566
3	444	517	581
4	454	525	583
5	459	534	600
6	473	546	612
7	474	540	615
8	493	562	610
9	507	561	629
10	517	590	627

Źródło: Obliczenia własne.

Dwustopniowe regresje grupowe wyznaczone na podstawie danych zawartych w tabelach 1 i 2 przedstawiają się następująco:

— dla grupy I:

$$\left(\frac{\hat{Y}_1}{\bar{Y}_1}\right) = 0{,}311\left(\frac{X_1}{\bar{X}_1}\right) + 420{,}7$$

— dla grupy II:

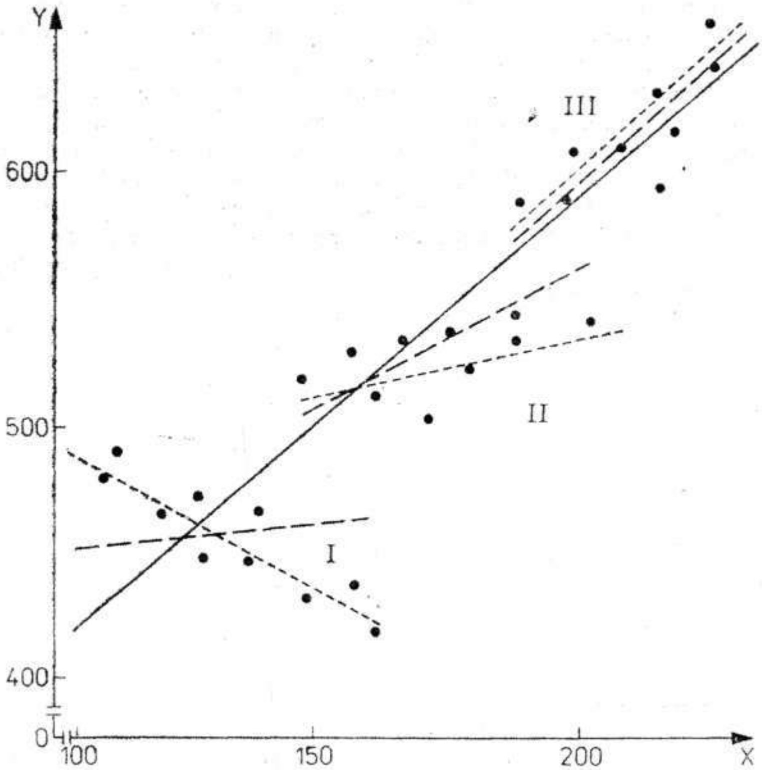
$$\left(\frac{\hat{Y}_2}{\bar{Y}_2}\right) = 1{,}053\left(\frac{X_2}{\bar{X}_2}\right) + 349{,}0,$$

— dla grupy III:

$$\left(\frac{\hat{Y}_3}{\bar{Y}_3}\right) = 1{,}739\left(\frac{X_3}{\bar{X}_3}\right) + 243{,}1.$$

Kształtowanie się dwustopniowych regresji grupowych przedstawione jest na rycinie.

Rozrzut punktów empirycznych na wykresie wskazuje na niejednorodność zbioru wyróżnionych zakładów produkcyjnych ze względu na poziom produkcji oraz wysokość zatrudnienia, a tym samym na koniecz-



Regresje produkcji względem zatrudnienia

Objaśnienia: ..... dane empiryczne, —— regresja ogólna, - - - - klasyczna regresja grupowa, — — — dwustopniowa regresja grupowa

ność jego podziału na grupy typologiczne i budowę funkcji regresji dla poszczególnych grup. Dwustopniowe regresje grupowe kształtują się pomiędzy regresją globalną a klasycznymi regresjami grupowymi, ich współczynniki kierunkowe przy zmiennej objaśnianej X przybierają wartości pośrednie między współczynnikami tych dwóch regresji. Przytoczony przykład potwierdza, że proponowane rozwiązanie uwzględnia zarówno ogólną tendencję zmian zmiennej objaśnianej, jak i prawidłowości właściwe dla poszczególnych grup typologicznych.

REGRESSIVE STUDIES ON EFFECTIVENESS OF PRODUCTION IN THE CONDITIONS OF HETEROGENITY OF THE SET OF OBJECTS

Summary

The article attempts at discriminating a set of objects in the regressive study- in situation of their heterogenity. The first stage consists of a division of the set into typological groups and next, of regressive modelling in the received sub-sets.

The following phase, which is in the focus of this study, involves proceeding with group regressive examinations. Suggested approaches to the regressive modelling, distinct to the classical one, is compromising between a general regression related to the whole set and a group regression which is designed in a traditional way i.e. a two phase regression. It proves correct to estimate a general regression for the whole set of objects and for groups and on these grounds — to create a two stage model which can explain changes of an explained variable in the respective groups which are resulting from the properties of the whole set of objects as well as of specific groups. The study is illustrated with the construction of one variable models on the example of simulational data for 30 plants isolated into three typological groups.