

*Daniel Kosiorowski\**

## ANALIZA DANYCH PANELOWYCH Z WYKORZYSTANIEM GŁĘBI REGRESYJNEJ

**Streszczenie.** W artykule proponuje się odporne podejście do estymacji parametrów liniowego modelu mieszanego dwóch zmiennych wykorzystujące koncepcję głębi regresyjnej. Wybrane własności proponowanego podejścia porównuje się z własnościami powszechnie wykorzystywanego uogólnionego estymatora najmniejszych kwadratów.

**Słowa kluczowe:** Głębia regresyjna, Dane panelowe, Liniowy model mieszany.

### I. WPROWADZENIE

W powszechnie wykorzystywanym w praktyce modelu liniowym pierwszego stopnia na ogół zakłada się, że obserwacje są pobierane z tej samej populacji, są niezależne i o takim samym rozkładzie (losowość pojawia się na poziomie błędu pomiaru). W przypadku analizy regresji prowadzonej z wykorzystaniem modeli mieszanych stosuje się mniej restrykcyjne założenia. Mianowicie dane mogą tworzyć skupiska. Zakładamy, że obserwacje pomiędzy skupiskami są niezależne jednak, nie muszą być niezależne wewnątrz skupisk. Modele, w których rozważa się jedynie ustalone poziomy zmiennych objaśniających nazywane są modelami z ustalonymi poziomami, modele, w których rozważa się jedynie losowe poziomy zmiennych objaśniających nazywa się modelami z poziomami losowymi<sup>1</sup>. W zastosowaniach ekonomicznych modele mieszane posiadają większe możliwości, ponieważ na ogół zjawiska ekonomiczne odznaczają się jednocześnie ustalonymi jak i losowymi efektami (niewyjaśniona wariancja, różne źródła wariancji). Modele mieszane są szczególnie odpowiednie w przypadku modelowania układów ekonomicznych wykazujących niejednorodność odpowiedzi na sygnał reprezentowany przez zmienne objaśniające.

Przypomnijmy, że tzw. klasyczny model prostej regresji dwóch zmiennych jest postaci:

---

\* Dr, Katedra Statystyki, Uniwersytet Ekonomiczny w Krakowie.

<sup>1</sup> W kontekście analizy danych panelowych wyróżnia się w literaturze ekonometrycznej (patrz Greene 2005) m. in. modele z ustalonymi poziomami (dopuszcza się specyficzne dla skupiska stałe), modele z charakterystycznymi dla skupiska efektami losowymi (dodatkowe źródło niepewności) nieskorelowanymi z regresorami, modele z losowymi parametrami, modele w których oczekiwana wartość parametru zależy od skupiska. Prezentowany w niniejszej pracy estymator nawiązuje do ostatniego przypadku.

$$y_k = \alpha + \beta x_k + \varepsilon_k, \quad k = 1, \dots, K, \quad (1)$$

gdzie  $\varepsilon_k$  są niezależnymi zmiennymi o identycznym rozkładzie o przeciętnej zero i takiej samej wariancji  $\sigma^2$ .

W przypadku **modelu mieszanego** na obserwacje patrzymy z punktu widzenia:

$$y_{ij} = \underbrace{\beta_0 + \beta_1 x_{ij}}_{ustalone} + \underbrace{\alpha_{0i} + \alpha_{1i} x_{ij}}_{losowe} + \underbrace{\varepsilon_{ij}}_{blad}, \quad i = 1, \dots, N, \quad j = 1, \dots, n_i, \quad (2)$$

gdzie:

$i$  – oznacza  $i$  te skupisko,

$n_i$  – liczba obserwacji w  $i$  – tym skupisku,

$\beta_0$  – wyraz wolny uśredniony dla całej populacji

$\beta_1$  – nachylenie uśrednione dla całej populacji

$\alpha_{0i}$  – efekt losowy dotyczący wyrazu wolnego w  $i$  tym skupisku

$\alpha_{1i} x_{ij}$  – efekt losowy dotyczący nachylenia w  $i$  tym skupisku

$\varepsilon_{ij}$  – niezależne błędy o takim samym rozkładzie i o tej samej wariancji  $\sigma^2$ .

Na ogół przyjmuje się, że  $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ ,  $\alpha_{0i} \sim N(m_{0i}, \sigma_0^2)$ ,

$\alpha_{1i} \sim N(m_{1i}, \sigma_1^2)$  są wzajemnie niezależne.

W najprostszym przypadku liniowego modelu mieszanego zakładamy, że losowe są jedynie wyrazy wolne. Przy tych samych oznaczeniach model przyjmuje wówczas postać:

$$y_{ij} = \underbrace{\beta_0 + \beta_1 x_{ij}}_{ustalone} + \underbrace{\alpha_{0i}}_{losowe} + \underbrace{\varepsilon_{ij}}_{blad}, \quad i = 1, \dots, N, \quad j = 1, \dots, n_i. \quad (3)$$

Modele mieszane wydają się być użyteczne np. w badaniach związków pomiędzy zmiennymi w ujęciu gmin z uwzględnieniem podziału na województwa, w badaniu dochodów i wydatków konsumpcyjnych z uwzględnieniem grupy dochodowej itd. Modele mieszane można stosować w przypadku **danych panelowych** to znaczy wówczas, gdy dane dotyczą poszczególnych jednostek przekrojowych w dłuższym czasie (więcej niż jednym okresie). Zauważmy, że zaletą danych panelowych jest możliwość złagodzenia założeń, które domyślnie są przyjmowane w analizie danych przekrojowych

Modele definiowane równaniami 2 i 3 są przypadkami szczególnym liniowego modelu mieszanego zaproponowanego przez Laird'a i Ware'a w 1982 roku:

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, N, \quad (4)$$

gdzie:

$\mathbf{y}_i$  – jest  $n_i \times 1$  wektorem zmiennej objaśnianej w  $i$  tym skupisku ( $n_i$  – odpowiedzi jednostek z  $i$  tego skupiska)

$\mathbf{x}_i$  – jest  $n_i \times m$  macierzą ustalonych efektów w  $i$  tym skupisku.

$\boldsymbol{\beta}$  – jest uśrednionym dla wszystkich skupisk wektorem parametrów związanych z ustalonymi efektami (z zależnością deterministyczną pomiędzy zmiennymi),

$\mathbf{Z}_i$  – jest  $n_i \times k$  macierzą eksperymentu efektów losowych w  $i$  tym skupisku (macierz złożona z zer i jedynek wskazująca, które ze współczynników przy zmiennych objaśniających traktujemy jako losowe).

$\boldsymbol{\varepsilon}_i$  –  $n_i \times 1$  wektor błędów dla  $i$  tego skupiska, wektor o niezależnych składowych, każda o przeciętnej zero i wariancji  $\sigma^2$ .

$\mathbf{b}_i$  – jest  $k \times 1$  wektorem parametrów związanych z efektami losowymi w  $i$  tym skupisku, wektor o zerowej przeciętnej i macierzy kowariancji  $\mathbf{D}_* = \sigma^2\mathbf{D}$ .

Zakładamy, że macierz  $\sum \mathbf{X}_i^t\mathbf{X}_i$  jest nieosobliwa oraz, że  $\sum n_i > m$  dla zapewnienia identyfikowalności modelu (6) względem  $\boldsymbol{\beta}$ . Dla zapewnienia identyfikowalności modelu (4) względem  $\sigma^2$  i  $D$ , zakładamy, że przynajmniej jedna macierz  $\mathbf{Z}_i^t\mathbf{Z}_i$  jest dodatnio określona oraz, że  $\sum_{i=1}^N (n_i - k) > 0$ .

Model 4 szacuje się najczęściej metodą największej wiarygodności przy założeniu, że  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$ ,  $\mathbf{b}_i \sim N(\mathbf{0}, \sigma^2\mathbf{D})$ . Przy takich założeniach model 4 można zapisać w następującej postaci brzegowej

$$\mathbf{y}_i \sim N\left(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2(\mathbf{I}_{n_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^t)\right), \quad i = 1, \dots, N. \quad (5)$$

Okazuje się, że ustalając macierz  $\mathbf{D}$ , logarytm wiarygodności dla modelu 4 maksymalizowany jest przez uogólniony estymator najmniejszych kwadratów:

$$\hat{\boldsymbol{\beta}}_{UNK} = \left[ \sum_{i=1}^N \mathbf{X}_i^t (\mathbf{I} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t)^{-1} \mathbf{X}_i \right]^{-1} \left[ \sum_{i=1}^N \mathbf{X}_i^t (\mathbf{I} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t)^{-1} \mathbf{y}_i \right]. \quad (6)$$

Zauważmy, że w specjalnym przypadku, gdy  $\mathbf{D} = \mathbf{0}$ , estymator (6) sprowadza się do zwykłego estymatora najmniejszych kwadratów

$$\hat{\boldsymbol{\beta}}_{NK} = \left( \sum \mathbf{X}_i^t \mathbf{X}_i \right)^{-1} \left( \sum \mathbf{X}_i^t \mathbf{y}_i \right). \quad (7)$$

Model definiowany równaniem (4) może pojawić się przykładowo w następującej sytuacji. Rozważamy grupę  $N$  przedsiębiorstw należących do różnych sektorów gospodarki. W przypadku każdego sektora badamy jak się ma zysk przypadający na pracownika względem płacy pracownika, jego stażu pracy, świadczeń socjalnych, efektywności pracy. Przypuśćmy, że interesuje nas płacy od pozostałych zmiennych zależność przyrostu naturalnego od pozostałych zmiennych. Kluczowym dla stosowalności modelu 4 w przedstawionej sytuacji jest założenie, że struktura zależności nie zmienia się od sektora do sektora co oznacza, że wektor nachyleń  $\boldsymbol{\beta}$  jest ustalonym wektorem.

Model z losowymi wyrazami wolnymi wydaje się bardziej realistyczny niż model klasyczny, gdyż dopuszcza charakterystyczny dla każdego z sektorów oddzielnie, poziom płacy.

Uogólniony Estymator NK podobnie jak zwykły estymator NK jest skrajnie nieodporny na jednostki odstające, punkt załamania wektora parametrów wynosi 0%.

Eksperymenty symulacyjne prowadzone przez autora z wykorzystaniem pakietu środowiska R {Imm} wskazują na niską efektywność estymatora UNK w przypadkach, gdy wariancja efektów losowych jest istotnie większa od wariancji błędu oraz gdy wariancja efektów losowych lub wariancja błędu są nieznane.

## II. PROPOZYCJA ODPORNEGO ESTYMATORA WYKORZYSTUJĄCEGO KONCEPCJĘ GŁĘBI REGRESYJNEJ

Brak odporności typowych estymatorów parametrów liniowego modelu mieszanego skłaniają autora do przedstawienia propozycji odpornego estymatora parametrów takiego modelu wykorzystującego koncepcję głębi regresyjnej Rouseeuw i Hubert (1998).

Zaznaczmy, że znane są w literaturze alternatywne propozycje odpornych estymatorów oraz testów odnoszących się do parametrów liniowego modelu mieszanego. Przykład stanowią prace Huggins i Staudte (1994) bądź Welsh i Richardson (1997) (ważona funkcja wiarygodności, problemy z odstawianiem co do zmiennych objaśniających, złożoność obliczeniowa),  $S$  – estymatory o wysokim punkcie załamania proponowane przez Copt i Victoria – Feser (2005) (rozszerzyli definicję wielowymiarowych  $S$ - estymatorów na przypadek liniowych modeli mieszanych), MM- estymator zaproponowany przez Copt i Hertier (2006). Ostatnie dwie propozycje zakładają wielowymiarową normalność rozpatrywanych układów. Założenie to często nie jest spełnione w przypadku układów ekonomicznych.

Powszechnie znany jest fakt, że metoda najmniejszych kwadratów jest rozszerzeniem na zagadnienie regresji średniej arytmetycznej. Rozszerzenie mediany Tukey’a na zagadnienie regresji zostało zaproponowane przez Rousseeuw i Hubert (1998). Rozważmy zbiór danych  $\mathbf{Z}^n$  o  $p-1$  zmiennych objaśniających  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p-1})$  i zmiennej objaśnianej  $y_i$  dla  $i = 1, \dots, n$ . W zagadnieniu regresji szukamy dopasowania postaci  $y = b_0 + b_1x_1 + \dots + b_{p-1}x_{p-1}$  dla pewnego  $\mathbf{b} = (b_0, \dots, b_{p-1}) \in \mathbb{R}^p$ , gdzie  $b_1, \dots, b_{p-1}$  to parametry nachylenia,  $b_0$  to wyraz wolny. **Głębia regresyjna** Rousseeuw i Hubert możliwego dopasowania  $\mathbf{b}$  wskazuje jak dobrze równanie regresji wyznaczone przez  $\mathbf{b}$  pasuje do danych. Autorzy definiują głębię regresyjną możliwego dopasowania  $\mathbf{b} \in \mathbb{R}^p$  względem próby  $(\mathbf{x}_i, y_i)$  jako

$$GR(\mathbf{b}, \mathbf{Z}^n) = 1/n \min_{\|\boldsymbol{\lambda}\|=1} \left\{ \frac{r_i(\mathbf{b})}{\mathbf{x}_i \boldsymbol{\lambda}^t} < 0, \mathbf{x}_i \boldsymbol{\lambda}^t \neq 0 \right\}, \quad (8)$$

gdzie  $\boldsymbol{\lambda} \in \mathbb{R}^p$ ,  $r_i(\mathbf{b})$  oznacza  $i$  – tą resztę regresji.

**Estymator maksymalnej głębi regresyjnej** definiujemy jako

$$MAXGR(\mathbf{b}, \mathbf{Z}^n) = \arg \max_{\mathbf{b}} GR(\mathbf{b}, \mathbf{Z}^n) \quad (9)$$

Estymator maksymalnej głębi regresyjnej Rousseeuw i Hubert odznacza się punktem załamania bliskim 30%, dobrą szybkością zbieżności oraz efektywnością. O modelu generującym dane zakładamy jedynie, że warunkowa mediana odpowiedzi jest liniowa względem zmiennych objaśniających. Estymator ten radzi sobie z heteroskedastycznością, autokorelacją i skośnością błędu.

**PROPOZYCJA:** Przypuśćmy, że dysponujemy próbą obserwacji par zmiennych objaśnianej i objaśniającej w podziale na skupiska jak w modelu definiowanym przez (3) Oznaczmy przez  $\mathbb{Z}_i^{n_i}$  zbiór par obserwacji  $y$  i  $x$  w  $i$  tym skupisku tzn.

$$\mathbb{Z}_i^{n_i} = \left( (y_{i1}, x_{i1})^t (y_{i2}, x_{i2})^t \dots (y_{in_i}, x_{in_i})^t \right)^t, \quad i = 1, \dots, N,$$

oraz oznaczmy przez  $\mathbb{Z}$  wszystkie obserwacje tzn.  $\mathbb{Z} = \mathbb{Z}_1^{n_1} \cup \dots \cup \mathbb{Z}_N^{n_N}$ .

Niech  $\hat{b}_i = DR(b, \mathbb{Z}_i^{n_i}) = (\hat{b}_i^0, \hat{b}_i^1)$  oznacza estymator maksymalnej głębi regresyjnej wyrazu wolnego i nachylenia w  $i$ -tym skupisku. Niech  $\hat{b} = DR(b, \mathbb{Z}^{n_i}) = (\hat{b}^0, \hat{b}^1)$  oznacza estymator maksymalnej głębi regresyjnej wyrazu wolnego i nachylenia dla wszystkich rozpatrywanych obserwacji. **Użytkujemy oszacowania prostego** liniowego modelu mieszanego jako:

$$\beta_1^{DR} = \hat{b}^1 - \text{nachylenie dla wszystkich rozpatrywanych skupisk},$$

$$\beta_0^{DR} = \hat{b}^0 - \text{wyraz wolny dla wszystkich rozpatrywanych skupisk},$$

$m_{0i}^{DR} = b_{0i}^{DR} - b_0^{DR}$  – efekt losowy w  $i$ -tym skupisku dotyczący wyrazu wolnego

$m_{1i}^{DR} = b_{1i}^{DR} - b_1^{DR}$  – efekt losowy w  $i$ -tym skupisku dotyczący nachylenia.

U podłoża powyższej propozycji leży prosty pomysł. Obliczamy mianowicie oszacowanie maksymalnej głębi regresyjnej dla wszystkich obserwacji traktowanych łącznie i oddzielnie dla każdego ze skupisk. Na bazie różnic pomiędzy tymi oszacowaniami obliczamy parametry liniowego modelu mieszanego.

### III. WYBRANE WŁASNOŚCI PROPONOWANEGO ESTYMATORA

W celu zbadania statystycznych własności proponowanego estymatora generowano 1000 prób złożonych z 60 obserwacji z punktu widzenia modelu

$$y_{ij} = 2 + \underbrace{b_i}_{a_i} + 10x_{ij} + \varepsilon_{ij}, \quad i = 1, 2, 3; \quad j = 1, \dots, 20, \quad (10)$$

gdzie  $\varepsilon_{ij} \sim N(0, 10)$ ;  $b_1 \sim N(2, 3)$ ,  $b_2 \sim N(5, 3)$ ,  $b_3 \sim N(0, 2)$ .

W celu oceny obciążenia i efektywności proponowanego estymatora maksymalnej głębi w porównaniu z uogólnionym estymatorem NK – dla każdej z 1000 prób oszacowano parametry modelu obiema metodami.

Tabela 1. Wyniki symulacyjnego badania nieobciążoności oraz rozrzutu oszacowań UNK parametrów liniowego modelu mieszanego

Ocena UNK	$\beta$	$\alpha$	$\alpha_1$	$\alpha_2$	$\alpha_3$
Przeciętna	9.99	2.3	1.86	3.13	1.02
Wariancja	0.0008	1.85	0.99	2.31	1.97

Źródło: Obliczenia własne.

Tabela 2. Wyniki symulacyjnego badania nieobciążoności oraz rozrzutu oszacowań RDEPTH parametrów liniowego modelu mieszanego

Ocena RDEPTH	$\beta$	$\alpha$	$\alpha_1$	$\alpha_2$	$\alpha_3$
Przeciętna	9.99	3.51	4.16	7.81	2.37
Wariancja	0.001	4.24	13.01	19.35	18.9

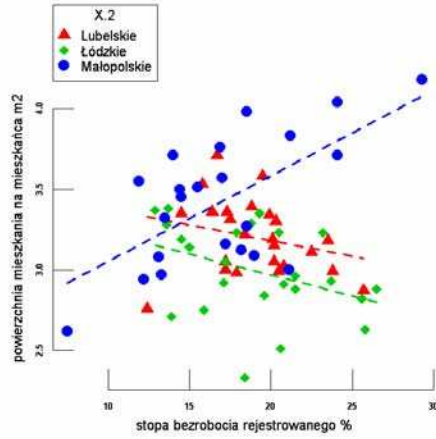
Źródło: Obliczenia własne

W tabelach 1 i 2 podsumowano wyniki symulacji w zakresie nieobciążoności i rozrzutu estymatorów UNK i maksymalnej głębi regresyjnej (RDEPTH). Z tabel wynika, że zarówno estymator UNK jak i RDEPTH są nieobciążonymi estymatorami nachylenia o porównywalnym rozrzucie. Z tabel wynika także, że estymator RDEPTH odznacza się mniejszym obciążeniem przy szacowaniu efektów losowych aniżeli estymator UNK.

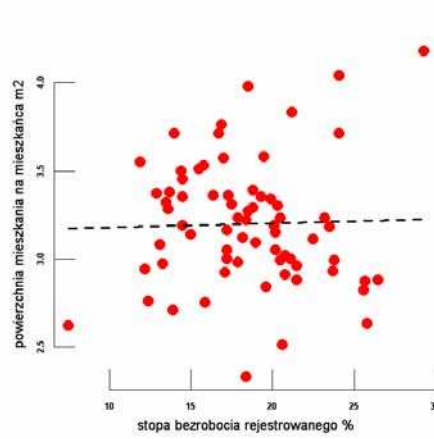
#### IV. PRZYKŁAD

Rozważmy zbiór danych złożony z 69 powiatów województw lubelskiego (24), łódzkiego (23) i małopolskiego (22) badanych ze względu na stopę bezrobocia rejestrowanego i powierzchnię mieszkania na 1 mieszkańca (w m<sup>2</sup>) w roku 2005 (dane GUS).

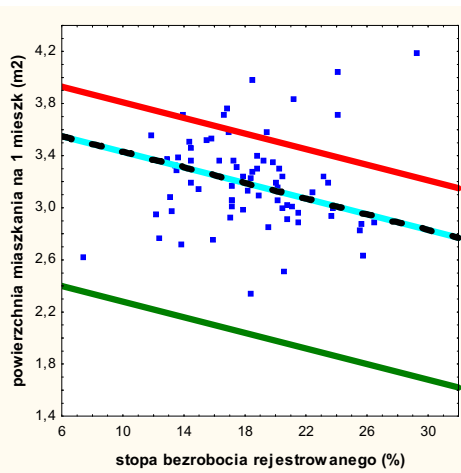
Na rys. 1 przedstawiono oszacowane modele regresji dla rozpatrywanych zmiennych dla powiatów w każdym województwie traktowanych oddzielnie, zauważmy różnice co do natury związku w województwie małopolskim oraz lubelskim i łódzkim. Na rys. 2 przedstawiono oszacowany model regresji dla rozpatrywanych zmiennych dla powiatów w trzech województwach traktowanych łącznie.



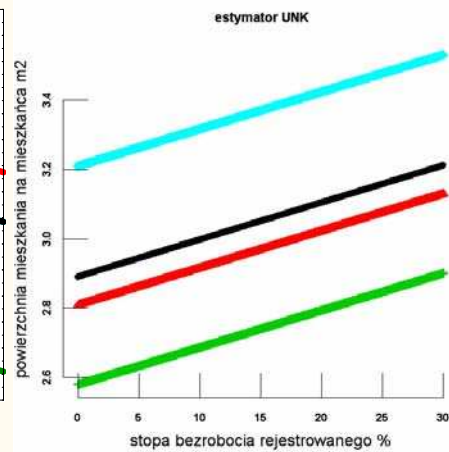
Rys. 1. Powierzchnia mieszkania vs. stopa bezrobocia – dane dotyczące powiatów trzech województw traktowanych oddzielnie.



Rys. 2. Powierzchnia mieszkania vs. stopa bezrobocia – dane dotyczące powiatów trzech województw traktowanych łącznie

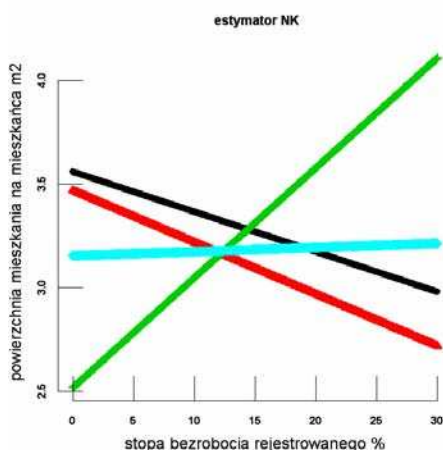


Rys. 3. Powierzchnia mieszkania vs. stopa bezrobocia – dane dotyczące powiatów w trzech województwach – estymator maksymalnej głębi parametrów modelu

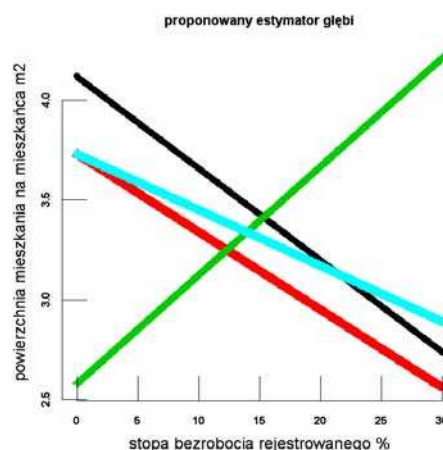


Rys. 4. Powierzchnia mieszkania vs. stopa bezrobocia – dane dotyczące powiatów w trzech województwach – estymator UNK parametrów modelu





Rys. 5. Powierzchnia mieszkania vs. stopa bezrobocia – dane dotyczące powiatów w trzech województwach – niezależne estymatory NK parametrów modelu



Rys. 6. Powierzchnia mieszkania vs. stopa bezrobocia – dane dotyczące powiatów w trzech województwach – niezależne estymatory maks. głębi parametrów modelu

Na rys. 3 przedstawiono oszacowany liniowy model mieszany z losowym wyrazem wolnym proponowaną metodą maksymalnej głębi regresyjnej. Na rys. 4 przedstawiono oszacowany liniowy model mieszany z losowym wyrazem wolnym metodą UNK. Natomiast na rys. 5 przedstawiono cztery niezależne modele liniowe ( dla każdego województwa i wszystkich danych traktowanych łącznie) szacowane metodą NK. Łatwo zauważyć, że proponowany estymator prowadzi do trafniejszej prezentacji zjawiska niż estymator UNK ponieważ wyraża tendencję specyficzną dla dwóch (większości obserwacji) województw (łódzkiego i lubelskiego). Na rys. 5 i 6 widzimy natomiast cztery niezależne modele liniowe wykonane metodą NK (rys. 5) i metodą głębi regresyjnej (rys. 6). Metoda głębi regresyjnej wykazuje przewagę także i w takim przypadku lepiej uwzględniając tendencję przeważającej części obserwacji.

## V. PODSUMOWANIE

Zaletą danych panelowych jest możliwość weryfikacji oraz złagodzenia założeń, które domyślnie są przyjmowane w analizie danych przekrojowych.

Przy stosunkowo łagodnych założeniach dotyczących mediany warunkowego błędu punkty załamania estymatora maksymalnej głębi w każdym ze skupisk wynoszą niezależnie  $BP \geq \frac{1}{d+1}$ , gdzie  $d$  oznacza liczbę zmiennych objaśniających.

Estymator maksymalnej głębi regresyjnej dobrze radzi sobie ze skośnymi oraz heteroskedastycznymi rozkładami błędów i efektów losowych.

Wyniki symulacji wskazują na nieobciążoność i zadowalającą „efektywność” estymatora maksymalnej głębi regresyjnej w porównaniu z uogólnionym estymatorem NK oraz niezależnymi ocenami parametrów w skupiskach za pomocą zwykłego estymatora NK.

#### BIBLIOGRAFIA

- Chatterjee, S., Hadi, A. S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, 1(3), s. 379–416.
- Copt, S., Hertier S. (2006). Robust MM-Estimation and Inference in Mixed Linear Models. <http://www.unige.ch/ses/metri/> Jan 2006.
- Copt, S., Victoria-Feser, (2006): High Breakdown Inference For Mixed Linear Models. *Journal of the American Statistical Association* 101, s. 292–300
- Demidenko, E. (2004). *Mixed Models – Theory and Applications*. Hoboken, New Jersey: John Wiley & Sons.
- Greene, W. H. (2005). *Econometric Analysis*, Prentice Hall, New York
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics: The approach based on influence functions*. Nowy York: John Wiley & Sons.
- Huber, P., Ronchetti, E. M. (2009). *Robust Statistics*. John Wiley & Sons. New York
- Huggins, R. M., Staudte R. G. (1994) Variance Components Models for Dependent Cell Populations. *Journal of the American Statistical Association* 89, s. 19–29.
- Kosiorowski, D. (2007). O Odpornej Analizie Regresji w Ekonomii na Przykładzie Koncepcji Głębi Regresyjnej. *Przegląd Statystyczny* (1), s. 109–121.
- Kosiorowski, D. (2008). Robust Classification and Clustering Based on the Projection Depth Function. P. Brito (Red.). (strony 209–216). *Physica – Verlag*, Heidelberg.
- Laird N., Ware J. (1982). Random-effects Models for Longitudinal Data, *Biometrics* 38, 963–974.
- McCulloch, Ch. E., Searle, S. R., Neuhaus, J. M. (2008). *Generalized, linear, and Mixed Models*. John Wiley & Sons, Hoboken, New Jersey.
- Rousseeuw, J. P., Hubert, M. (1998). Regression Depth. *Journal of The American Statistical Association* (94), s. 388–433.
- Welsh, A.H., Richardson, A.M., (1997). Approaches to the Robust Estimation of Mixed Models *Handbook of Statistics Vol 15*, Elsevier Science B.V.

*Daniel Kosiorowski*

#### REGRESSION DEPTH BASED ANALYSIS OF PANEL DATA

##### Abstract

In this paper a robust approach to the linear mixed model parameters estimation is proposed. The approach appeals to the regression depth concept. Selected statistical features of the proposition in a comparison to a generalized least squares estimator are investigated using Monte Carlo approach.