Górnictwo i Geoinżynieria • Rok 30 • Zeszyt 3/1 • 2006

*Dariusz Foszcz**

ESTYMACJA PARAMETRÓW FUNKCJI REGRESJI METODĄ KLASYCZNĄ ORAZ METODAMI BOOTSTRAPOWYMI**

1. Wstęp

W zmieniającej się rzeczywistości przebiegu procesów technologicznych coraz częściej klasyczne metody statystyczne nie wystarczają do precyzyjnego opracowania i prezentowania wyników różnego typu badań. Konieczne staje się korzystanie z metod, które odchodzą od klasycznych założeń modelu statystycznego, ograniczających zakres ich zastosowania. Sięgać musimy zatem do nieklasycznych metod statystycznych a, w szczególności do wnioskowania nieparametrycznego (nieparametrycznej estymacji parametrów i charakterystyk funkcyjnych rozkładu czy też testów nieparametrycznych), które z jednej strony, pozwala na weryfikację założeń metod parametrycznych, z drugiej natomiast, przy niespełnionych założeniach ułatwia znalezienie wielu innych możliwych podejść. Coraz szersze zastosowanie metod statystycznych w opisie i ocenie procesów technologicznych wynika z rozwoju statystyki oraz możliwości obliczeniowych komputerów i opracowywanych programów statystycznych. Tworzone są programy od uniwersalnych pakietów statystycznych po programy specjalistyczne. Potwierdzeniem tego jest zauważalny postęp w rozwoju teorii tych metod. Nie do pominięcia jest konieczność stosowania statystyki w kontroli jakości. Opracowywane nowe normy jakości oraz zrewidowane stare normy zawierają wskazówki jakie narzędzia i techniki należy i można stosować do zrealizowania konkretnych zadań projakościowych.

W przeróbce surowców mineralnych metody statystyczne stosuje się przy analizie efektów prac eksperymentalnych, których głównymi celami są ocena wzbogacalności surowca i możliwości jej poprawy oraz informacje o przebiegu procesów przeróbczych pod kątem określenia czynników, dających najlepsze rezultaty przy danym sposobie wzbogacania. Metody statystyczne stosowane są również w badaniach mechanizmu zjawisk zacho-

* Wydział Górnictwa i Geoinżynierii, Akademia Górniczo-Hutnicza, Kraków

** Artykuł jest wynikiem realizacji pracy statutowej AGH nr 11.11.100.238

dzących w procesach przeróbczych oraz w poszukiwaniach optymalnych warunków przebiegu procesu. Opracowanie wyników eksperymentu czynnego lub biernego odbywa się poprzez zastosowanie odpowiednich metod statystycznych [2].

Optymalnie dobrane metody statystyczne stosowane w tych zagadnieniach zostały adoptowane adekwatnie do potrzeb. Stosowane metody posiadają niestety pewne wady. Do głównych wad i utrudnień w stosowaniu tych metod, które można określić terminem „klasycznych", należy zaliczyć występujące założenia ograniczające możliwości ich zastosowania. Stąd też pojawiło się szerokie zainteresowanie testami pozwalającymi na sprawdzenie spełnienia takich założeń w danym badaniu empirycznym lub tworzeniem innych metod wnioskowania statystycznego, w których założenia te nie występują. Coraz większe zapotrzebowanie na metody statystyczne charakteryzujące się stosunkowo nielicznymi i niekłopotliwymi założeniami ograniczającymi zakres ich stosowania spowodowało powstanie metod określanych mianem nieklasycznych metod statystycznych.

Podstawą niniejszego artykułu jest próba adaptacji nieklasycznych metod statystycznych (bootstrapowych) w problematyce estymacji parametrów modeli stochastycznych. Przeróbka surowców mineralnych — inżynieria mineralna, tak jak każda inżynieria techniczna podlega ciągłemu rozwojowi i doskonaleniu swoich technologii. Rozwijające się bardzo intensywnie techniki konstrukcji urządzeń kontrolno-pomiarowych oraz wchodząca do przemysłu komputerowa technika zbierania i przetwarzania danych tworzy dobre podstawy do doskonalenia modeli matematycznych przemysłowych układów technologicznych wzbogacania różnych surowców.

Modelowanie matematyczne procesów przemysłowych rozpoczęto od równań regresji, które oparte były o założenie, że w okolicach tzw. punktu pracy, zależności pomiędzy wskaźnikami oceny przebiegu procesu a wielkościami (zmiennymi technologicznymi) wpływającymi na jego wartość mogą być linearyzowane, lub generalnie, przedstawione w postaci wielomianu. Następnym etapem były próby tworzenia postaci regresyjnych modeli w oparciu o pewne przesłanki teoretyczne oraz zwrócenie uwagi na organizację danych w czasie (opóźnienia transportowe i masowe), czyli podzielono modele na statyczne i dynamiczne. To zdecydowało o stosowaniu korelacyjnej teorii procesów stochastycznych oraz teorii procesów Markowa. Następnym etapem stosowania modeli matematycznych w przeróbce były modele dyskretne uwzględniające krok próbkowania i opóźnienia czasowe. Ogólnie biorąc, są to modele typu ARMAX uwzględniające człony związane ze średnią ruchomą (MA) oraz człony związane z wielkościami egzogennymi (X) w różnych kombinacjach.

2. Podstawy teoretyczne

Metody analizy regresji używane są najczęściej w Statystyce na potrzeby opisu kształtowania się poziomu pewnego zjawiska w czasie, jak i na podstawie pobieranych z populacji generalnej prób losowych [3].

Jeżeli pomiędzy dwoma zmiennymi losowymi (cechami statystycznymi ilościowymi) istnieje zależność korelacyjna i jedną ze zmiennych y możemy uznać za zależną, a drugą x za niezależną, to można próbować sformułować zależność funkcyjną, która przedstawiałaby wartość y w zależności od wartości x i pewnej dodatkowej zmiennej losowej ξ, która reprezentuje losową zmienność zmiennej y i jest niezależna od x

$$y = f(x, \xi)$$

gdzie:

y — zmienna objaśniana (zależna),
x — zmienna objaśniająca (niezależna),
ξ — składnik losowy,
f — postać funkcji zależności.

Jeżeli wartość zmiennej y zależy od wartości wielu zmiennych niezależnych $x_1, x_2, ..., x_k$, mówimy o regresji wielowymiarowej (wielorakiej). Regresja wielowymiarowa dotyczy badania wpływu wielu zmiennych objaśniających na zmienną objaśnianą. Badanie takie może dotyczyć zmiennych losowych o nieznanym, jednak zakładanym rozkładzie normalnym lub zmiennych których wartości pochodzą z szeregów czasowych. W obu przypadkach możliwe jest przeprowadzanie analizy regresji. Jak można się spodziewać ten rodzaj analizy jest trudniejszy niż w przypadku regresji jednowymiarowej, bowiem w przypadku kilku już zmiennych objaśniających szybkich obliczeń można dokonywać jedynie przy pomocy komputerów. Model w przypadku regresji wielowymiarowej można zapisać w postaci

$$y = f(x_1, x_2, ..., x_k, \xi)$$

W przypadku, gdy funkcja f z powyższej zależności jest funkcją liniową, model przyjmuje postać

$$y = \alpha_0 + \alpha_1 x_1 + ... + \alpha_k x_k + \xi$$

gdzie:

y — zmienna objaśniana,
$x_1, x_2, ..., x_k$ — zmienne objaśniające,
ξ — składnik losowy,
$\alpha_1, \alpha_2, ..., \alpha_k$— wartości parametrów funkcji regresji (parametry modelu).

Wartości parametrów funkcji regresji na ogół nie są znane i do ich dokładnego wyznaczenia potrzebna byłaby znajomość rozkładów wartości zmiennej y dla wszystkich możliwych zestawów wartości zmiennych $x_1, x_2, ..., x_k$.

Model regresji można także zapisać w postaci macierzowej jako

$$\mathbf{y} = \mathbf{X}\alpha + \xi$$

gdzie:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$ — wektor wartości (realizacji) zmiennej objaśnianej,

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_k \end{bmatrix}$$ — wektor wartości parametrów modelu,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$ — macierz wartości zmiennych objaśniających,

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \dots \\ \xi_n \end{bmatrix}$$ — wektor wartości składnika losowego.

W ogólnym przypadku k zmiennych dla modelu regresji liniowej w postaci macierzowej

$$\mathbf{y} = \mathbf{Xa} + \mathbf{e} \text{ oraz } \hat{\mathbf{y}} = \mathbf{Xa}$$

z metody najmniejszych kwadratów wynika, że oceny parametrów modelu (realizacje estymatorów w próbie) można wyznaczyć jako

$$\mathbf{a} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Jeżeli spełnione są założenia modelu regresji liniowej, to estymatory parametrów równania regresji otrzymane metodą najmniejszych kwadratów (MNK-estymatory) są zgodne, nieobciążone i najbardziej efektywne w klasie estymatorów liniowych (Twierdzenie Gaussa-Markowa).

Dla jednej zmiennej objaśniające wzory macierzowe przyjmują postać:

$$\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Miarą przeciętnej wielkości błędu dopasowania jest wariancja resztowa, która jest oceną wariancji składnika losowego

$$S_e^2 = \frac{1}{n-(k+1)}\mathbf{e}^{\mathrm{T}}\mathbf{e} = \frac{1}{n-(k+1)}\sum_{i=1}^{n} e_i^2,$$

natomiast przeciętny błąd szacunku parametru *j* jest równy

$$S(a_j) = \sqrt{S_e^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}_{j+1,j+1}}$$

Elementy pod pierwiastkiem są kolejnymi elementami głównej przekątnej tzw. macierzy kowariancji (macierzy wariancji i kowariancji) ocen parametrów

$$\mathbf{D}^2(\mathbf{a}) = S_e^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

Średni względny błąd szacunku parametru *j* wyraża się wzorem

$$\frac{S(a_j)}{|a_j|}$$

Bardzo ważną częścią analizy regresji, po oszacowaniu ocen numerycznych parametrów modelu jest ocena zmienności zmiennej objaśnianej Y spowodowanej zmiennością zmiennej objaśniającej X. Do oceny takiej służą współczynniki determinacji i indeterminacji, których suma musi wynosić 1. Posługując się wzorem na wskaźnik indeterminacji, inaczej zwany wskaźnikiem zbieżności, którego wartość powinna być jak najniższa. Wskaźnik zbieżności ponadto, określa wartość współczynnika determinacji, który dany jest wzorem poniższym

$$R^2 = 1 - \varphi$$

Wskaźnik ten informuje jaki procent zmienności zmiennej objaśnianej wyjaśniony został zmiennością zmiennej objaśniającej.

W przypadku regresji liniowej, także wielorakiej należy przeprowadzić jeszcze kilka analiz wchodzących w skład analizy reszt regresyjnych potwierdzających przydatność skonstruowanego modelu w celach prognostycznych.

Współczynnik determinacji dany jest w tym przypadku wzorem

$$R^2 = 1 - \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{\mathbf{y}^{\mathrm{T}}\mathbf{y} - n(\overline{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

(określa, jaka część zmienności cechy zależnej jest wyjaśniona zmiennością cechy niezależnej).

Oszacowane parametry oraz skonstruowany model pozostaje jeszcze sprawdzić pod względem użyteczności bowiem nie każdy model nadaje się do dalszego za jego pomocą wnioskowania statystycznego. Jak interpretować oszacowane parametry modelu oraz, czy są one statystycznie istotne. Oczywistym jest że najpierw należy odpowiedzieć na to drugie pytanie, inaczej nie będzie sensu udzielania odpowiedzi na pytanie pierwsze.

3. Bootstrapowa estymacja parametrów funkcji regresji

Do estymacji parametrów funkcji regresji poza opisanymi już szeroko metodami klasycznymi możemy wykorzystać metody nieklasyczne tj. metody bootstrapowe [4].

Bootstrapowa estymacja punktowa parametrów funkcji regresji oparta na estymatorach uzyskiwanych metodą najmniejszych kwadratów (MNK).

Niech dany będzie model postaci

$$y_i = f\left(X_i, \beta\right) + \xi_i \text{ dla } i = 1, \ldots, n$$

gdzie:

y_i — realizacje zmiennej objaśnianej (zależnej),
X_i — realizacja wektora zmiennych objaśniających (niezależnych),
ξ_i — wartość składnika losowego dla $i = 1, \ldots, n$,
β — wektor nieznanych parametrów.

Model ten opisuje zależność między zmienną Y i wektorem zmiennych objaśniających X.

Niech $\hat{\beta}$ będzie MNK-estymatorem parametru β lub wartością tego estymatora. Oszacowania wartości składników losowych są wówczas określone wzorem

$$\hat{\xi}_i = y_i - f\left(X_i, \hat{\beta}\right) \text{ dla } i = 1, \ldots, n$$

Zdefiniujemy rozkład postaci

$$P\left(Z = \hat{\xi}_i\right) = \frac{1}{n} \text{ dla } i = 1, \ldots, n$$

Według tego rozkładu generujemy próbę bootstrapową $\left(\hat{\xi}_1^*, \ \ldots, \hat{\xi}_n^*\right)$, a następnie tworzymy próbę $\left(Y_1^*, \ \ldots, Y_n^*\right)$, której realizacjami są wartości $\left(y_1^*, \ \ldots, y_n^*\right)$ określone wzorem

$$y_i^* = f\left(X_i, \hat{\beta}\right) + \xi_i^* \text{ dla } i = 1, \ldots, n$$

Wartości $y_1^*, \ \ldots, y_n^*$ wykorzystujemy do ponownego oszacowania metodą MNK parametru β.

Estymację przeprowadzamy dla modelu

$$y_i^* = f\left(X_i, \beta\right) + \tilde{\xi}_i \text{ dla } i = 1, ..., n$$

gdzie $\tilde{\xi}_i$ jest składnikiem losowym w utworzonym modelu.

Przedstawiony schemat generowania ciągów $\left(\hat{\xi}_1^*, ..., \hat{\xi}_n^*\right)$ i wyznaczania oszacowania parametru β powtarzamy N razy. W wyniku otrzymujemy wartości $\left(\beta_1^*, ..., \beta_N^*\right)$, które określają bootstrapowy rozkład MNK-estymatora $\hat{\beta}$. Na ich podstawie można wyznaczyć histogram, który graficznie przedstawi ten rozkład.

Jeżeli przez $\hat{\beta}^*$ oznaczymy estymator bootstrapowy parametru β, to można wykazać, że charakteryzuje się on następującymi własnościami

$$E\left(\hat{\beta}^*\right) = \hat{\beta} \text{ i } \operatorname{cov}\left(\hat{\beta}^*\right) = \hat{\sigma}^2 \left(X^T X\right)^{-1}$$

gdzie $\hat{\beta}$ jest wartością MNK-estymatora parametru β, $X^T = \left(X_{1,} ..., X_n\right)$

oraz $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left[y_i - f\left(X_i, \hat{\beta}\right) \right]^2$.

4. Oszacowanie funkcji regresji metodą klasyczną i bootstrapową

W celu oceny możliwości wykorzystania nieklasycznych metod statystycznych w problematyce estymacji parametrów modeli stochastycznych przeprowadzono obliczenia zależności funkcyjnych pomiędzy uzyskiem i zawartością miedzi w nadawie. Obliczenia przeprowadzono dla średnich miesięcznych wyników technologicznych za rok 2005 z Oddziału Zakłady Wzbogacania Rud KGHM Polska Miedź SA Rejon Polkowice oraz Lubin. Estymację parametrów powyższej zależności metodą klasyczną MNK wykonano wykorzystując pakiet statystyczny STATISTICA 6.0, zaś obliczenia metodyką bootstrap przy użyciu arkusza kalkulacyjnego.

Na podstawie tych danych metodą najmniejszych kwadratów, a następnie metodą bootstrapową dokonywano oszacowania parametrów modelu

$$y_t = a_0 + a_1 x_t + \xi_t$$

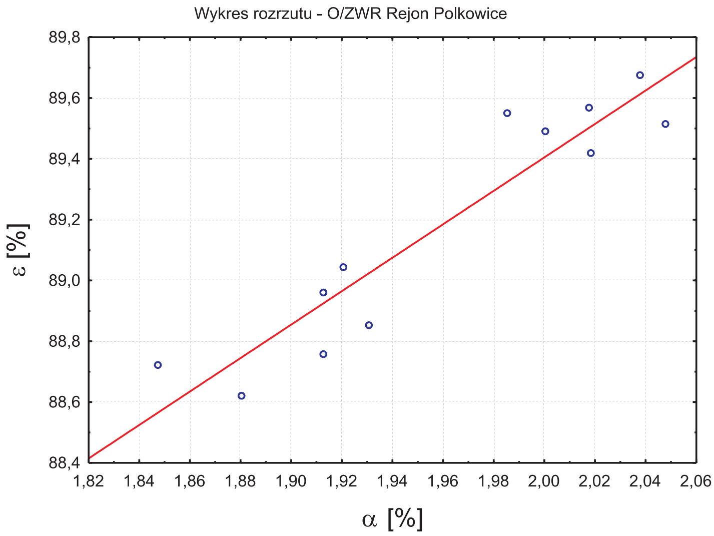
gdzie:

y_t — uzysk miedzi — ε,
x_t — zawartość miedzi w nadawie — α,
a_0, a_1 — oszacowywane parametrów modelu,
ξ_t — składnik losowy.

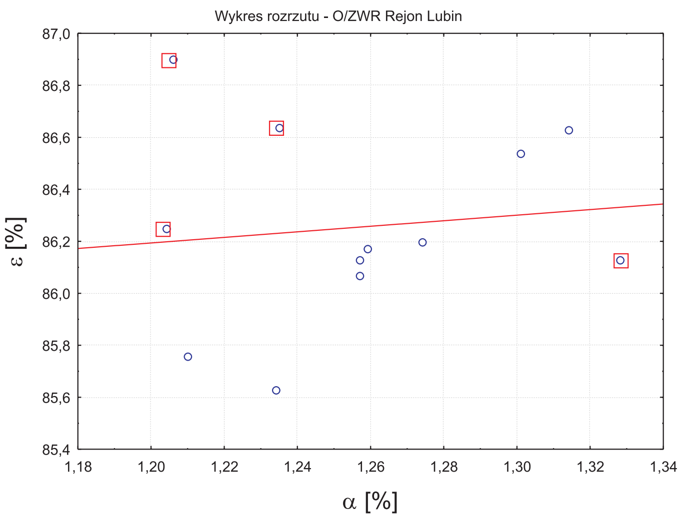
Estymacja parametrów funkcji regresji metodą najmniejszych kwadratów

Estymację parametrów zależności uzysku od zawartości miedzi w nadawie metodą klasyczną MNK wykonano wykorzystując pakiet statystyczny STATISTICA 6.0. Wykresy zależności zmiennej uzysku od zawartości miedzi w nadawie dla O/ZWR Rejon Polkowice oraz Lubin przedstawiono na rysunku 1. Są one pomocne w wyznaczaniu zależności funkcyjnych pomiędzy zmiennymi umożliwiając zorientowanie się o charakterze tej zależności [1].

a)



b)



Rys. 1. Zależność uzysku miedzi w koncentracie od zawartości miedzi w nadawie dla: a) O/ZWR Rejon Polkowice; b) O/ZWR Rejon Lubin

W tabeli 1 przedstawiono wyniki oszacowania dla parametrów równania zależności uzysku od zawartości miedzi w nadawie metodą klasyczną MNK.

TABELA 1

Wyniki estymacji parametrów zależności uzysku od zawartości miedzi w nadawie metodą klasyczną MNK

O/ZWR Rejon Polkowice	Podsumowanie regresji zmiennej zależnej: Uzysk ε $R^2 = 0{,}8674$, $F(1,10) = 72{,}973$, $p < 0{,}00001$, $S_r = 0{,}14218$				
		B	błąd standardowy	*t*(10)	poziom *p*
	a_0 — wyraz wolny	78,404	1,26	62,1137	0,000000
	a_1	5,500	0,64	8,5424	0,000007
O/ZWR Rejon Lubin	Podsumowanie regresji zmiennej zależnej: Uzysk ε $R^2 = 0{,}0146$, $F(1,10) = 0{,}14831$, $p < 0{,}70822$, $S_r = 0{,}38414$				
		B	błąd standardowy	*t*(10)	poziom *p*
	a_0 — wyraz wolny	84,911	3,49	24,3298	0,00000
	a_1	1,069	2,78	0,3851	0,70823

Uzyskano więc metodą MNK następującą postać funkcji regresji:

— O/ZWR Rejon Polkowice

$$y_t = 78{,}404 + 5{,}500 x_t$$

lub

$$\varepsilon = 78{,}404 + 5{,}500\alpha$$

— O/ZWR Rejon Lubin

$$y_t = 84{,}911 + 1{,}069 x_t$$

lub

$$\varepsilon = 84{,}911 + 1{,}069\alpha$$

Oszacowane zależność cechują się dla O/ZWR Rejon Polkowice wysoką wartością R^2 (86,74%) oraz istotnością współczynnika przy zmiennej niezależnej (zawartość miedzi w nadawie — α) a_1 a także dużą dokładnością z uwagi na niską wartość średniego odchylenia przeciętnego $S_r = 0{,}14$. Dla O/ZWR Rejon Lubin nieistotność współczynnika przy zmiennej niezależnej a_1 oraz bardzo niska wartość współczynnika determinacji R^2, jedynie średnie odchylenie przeciętne S_r równe 0,38 możemy uznać za zadowalające (należy jednak pamiętać że ta wartość średniego odchylenia przeciętnego wynika z niewielkiego zakresu zmienności uzysku miedzi — rys. 1b.). Uzyskane wartości szacowanego równania dla O/ZWR Rejon Lubin powodują że należy je uznać za nieistotne. Przyczyną są tzw. odstające wartości które zostały zaznaczone na rysunku 1b dla których uzyskano przy niskich wartościach zawartości miedzi bardzo rozbieżne wartości uzysku. Powodem tych rozbieżności poza błędami oznaczeń, które należy z uwagi na stosowane procedury wykluczyć zmiana wzbogacalności przerabianej rudy na skutek zmiany składu mineralogicznego. Z uwagi na fakt iż nie samo zagadnienie uzyskania poprawnego modelu było przedmiotem niniejszych rozważań przeprowadzono dalsze obliczenia oszacowania parametrów modelu metodą bootstrapową.

Bootstrapowa estymacja parametrów funkcji regresji

Do oszacowania parametrów modelu metodą bootstrapową zastosowano opisaną procedurę, przyjmując liczbę szacowań $N = 1020$. Uzyskane rezultaty przedstawiono w tabeli 2.

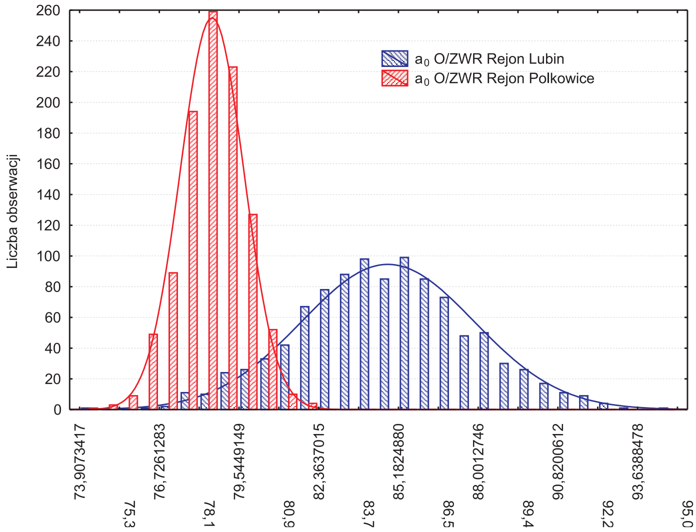
TABELA 2

Statystyka oszacowanych parametrów modelu metodą bootstrapową

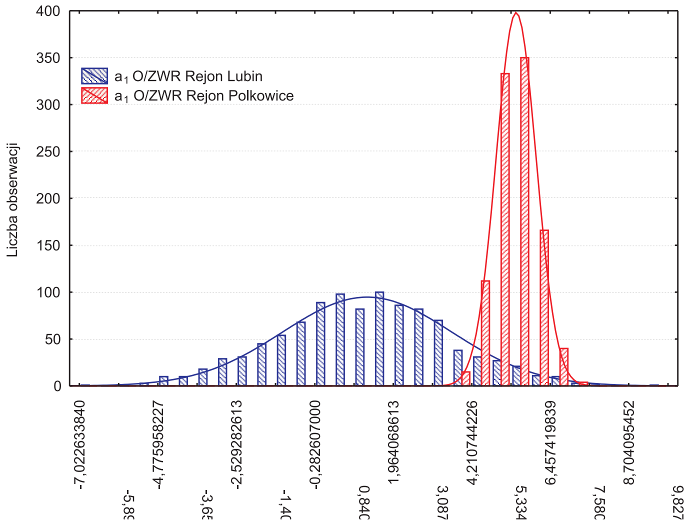
		n ważnych	Średnia	Minimum	Maksimum	Odchylenie standardowe
O/ZWR Rejon Polkowice	a_0 — wyraz wolny	1020	78,566	74,53	81,79	1,12
	a_1	1020	5,417	3,78	7,47	0,57
O/ZWR Rejon Lubin	a_0 — wyraz wolny	1020	84,755	73,91	95,05	3,03
	a_1	1020	1,175	–7,02	9,83	2,41

Na rysunkach 2 i 3 przedstawiono zaś histogramy rozkładów tych parametrów uzyskane na podstawie 1020 bootstrapowych repetycji.

Oszacowanie parametrów modelu przy pomocy metody bootstrap umożliwia szerszą analizę porównawczą uzyskanych wyników. Wynika z niej że dla równania O/ZWR Rejon Polkowice uzyskano mniejszą zmienność dla współczynników a_1 i a_0. Z uwagi na brak istotności uzyskanego równania dla O/ZWR Rejon Lubin nie ma możliwości jego interpretacji z uwagi na prowadzony proces technologiczny a w szczególności porównania z wartościami uzyskanymi dla O/ZWR Rejon Polkowice. Uzyskanie poprawnej zależności funkcyjnej pomiędzy badanymi zmiennymi umożliwia ocenę wpływu wzrostu zawartości miedzi w nadawie na wartość uzysku. Z równania uzyskanego dla O/ZWR Rejon Polkowice wynika że wzrost zawartości miedzi w nadawie o 0,1% powoduje wzrost uzysku o 0,54%. Oczywiście na wartość uzysku ma wpływ w głównej mierze poprawnie prowadzony proces wzbogacania, jednak z uzyskanego równania płyną pewne korzyści praktyczne np. możliwość porównania uzyskanych wartości uzysku przy zmiennej zawartości miedzi w nadawie. Ma to miejsce w sytuacji wprowadzania zmian w prowadzonej technologii czy też zastosowania nowych urządzeń i ocenie o ile udało się dzięki temu poprawić skuteczność prowadzonego procesu co mierzymy między innymi uzyskiem, którego wartość zgodnie ze wzorem obliczamy przy pomocy właśnie zawartości składnika użytecznego w nadawie.



Rys. 2. Histogramy rozkładu MNK-estymatora parametru a_0 modelu $y_t = a_0 + a_1 x_t + \varepsilon_t$ dla O/ZWR Rejon Polkowice i O/ZWR Rejon Lubin



Rys. 3. Histogramy rozkładu MNK-estymatora parametru a_1 modelu $y_t = a_0 + a_1 x_t + \varepsilon_t$ dla O/ZWR Rejon Polkowice i O/ZWR Rejon Lubin

5. Wnioski

Nieklasyczne metody statystyczne, a w szczególności wnioskowanie nieparametryczne (nieparametryczna estymacja parametrów i charakterystyk funkcyjnych rozkładu czy też testów nieparametrycznych), pozwalają na weryfikację założeń metod parametrycznych, a dodatkowo przy niespełnionych założeniach ułatwiają znalezienie wielu innych możliwych podejść.

Metody bootstrapowe pozwalają właśnie wyznaczyć, nie tylko oszacowania parametrów (w tym ich przedziały ufności), ale także aproksymację rozkładów estymatorów zastosowanych do początkowej oceny tych parametrów np. tak jak w analizowanym przypadku rozkładu MNK-estymatorów parametrów rozpatrywanej funkcji regresji. Możliwa staję się więc szersza analiza uzyskanych wyników wyznaczonej zależności funkcyjnej dla badanych parametrów. W analizowanym przypadku może to być analiza zmienności uzyskanej zależności pomiędzy uzyskiem a zawartością miedzi w nadawie dla okresów miesięcznych czy też pomiędzy poszczególnymi Rejonami Oddziału Zakłady Wzbogacania Rud KGHM Polska Miedź SA.

Problem wykorzystania metod nieklasycznych w tym bootstrapowych w ocenie zależności funkcyjnych parametrów, wymaga bardziej szczegółowej analizy teoretycznej oraz rachunkowej na większym materiale obserwacyjnym co będzie przedmiotem dalszych prac.

LITERATURA

[1] STATISTICA PL dla Windows. Ogólne konwencje i statystyki. T. 1, StatSoft 1997

[2] *Tumidajski T.*: Zastosowanie metod statystycznych w analizie procesów przeróbki surowców mineralnych. Katowice, Śląskie Wyd. Techn. 1993

[3] *Gajek L., Kałuszka M.*: Wnioskowanie statystyczne. Warszawa, WNT 1994

[4] *Domański Cz., Pruska K.*: Nieklasyczne metody statystyczne. Warszawa, PWE 2000