Ekonomia Menedżerska
2007, nr 1, s. 165–174

Anna Czapkiewicz*

Przykłady zależności pomiędzy dochodem a wydatkami na konsumpcję w przypadku losowości zmiennej niezależnej

1. Wstęp

W badaniach zależności wydatków na konkretne dobra i usługi od dochodu wykorzystuje się najczęściej model regresji liniowej lub funkcję potęgową, tj. funkcję, którą cechuje stała elastyczność. Modele takie buduje się w celu prognozowania wydatków w zależności od dochodu. Nieznane parametry występujące w proponowanych zależnościach szacuje się metodą najmniejszych kwadratów. Podejście takie wymaga założenia, że zmienna niezależna – dochód – jest nielosowa.

Podstawą znanej teorii konsumpcji Friedmana [5] jest model prostej regresji z błędami pomiaru obu zmiennych: zależnej i niezależnej. Zmienna zależna reprezentuje permanentną konsumpcję, natomiast zmienna niezależna – permanentny dochód. Podejście takie nieco się różni od modelu prostej regresji liniowej, w której zmienna niezależna jest deterministyczna.

Problem oszacowania parametrów zależności liniowej, gdy obserwowane cechy są zmiennymi losowymi jest znany w literaturze statystycznej. Przegląd takich modeli można spotkać np. w monografiach takich autorów jak A.W. Fuller [6], O. Bunke i H. Bunke [1].

Okazuje się, że naturalne uogólnienie schematu Gaussa–Markowa na przypadek losowych zmiennych objaśniających nie prowadzi do uzyskania zgodnych estymatorów nieznanych parametrów badanej zależności. Model taki jest nieidentyfikowalny, co oznacza, że nie istnieją metody, które wyznaczają co najmniej zgodne estymatory nieznanych parametrów regresji. Dopiero dodatkowe założenia

* Wydział Zarządzania, Akademia Górniczo-Hutnicza w Krakowie

o parametrach rozkładu zmiennych pozwalają rozwiązać problem nieidentyfikowalności.

Nałożenie dodatkowych warunków na parametry rozkładu zaburzeń nie jest satysfakcjonujące z punktu widzenia praktycznych zastosowań, ponieważ często parametrów rozkładu nie znamy, a tylko możemy oszacować je z próby. Okazuje się, że replikacja zmiennych losowych pozwala rozwiązać problem nieidentyfikowalności. Model z powtórzeniami i jego pewne własności przedstawione są m.in. w literaturze przedmiotu [3, 4].

Niniejsza praca stanowi próbę porównania dwóch metod w badaniu zależności liniowej w wypadku tych samych zmiennych losowych. Różnica polegać będzie na sposobie, w jaki zmienna niezależna będzie definiowana. Zamieszczony przykład będzie dotyczyć zależności miesięcznych wydatków na żywność i napoje bezalkoholowe od miesięcznego przychodu na osobę. Dane pochodzą z biuletynu GUS-u „Budżety gospodarstw domowych” od roku 1995.

2. Model funkcjonalny

Przed przystąpieniem do analizy przykładu przedstawiono pewne własności modelu, który będzie wykorzystany w opracowaniu:

Niech

$$X_{ij} = \xi_{ij} + \varepsilon_{ij} ,$$

$$Y_{ij} = a\xi_{ij} + b + \delta_{ij} ,$$

gdzie:

$i = 1,\ldots,n$,
$j = 1,\ldots,m$;

oraz

$$\varepsilon_{ij} \approx N(0,\sigma_\varepsilon^2) ,$$

$$\delta_{ij} \approx N(0,\sigma_\delta^2) ,$$

$$\xi_{ij} \approx N(s_i,\sigma_s^2) .$$

Zakładamy, że wektor (X_{ij},Y_{ij}) ma rozkład normalny z wartością oczekiwaną $(s_i, as_i + b)$ i macierzą kowariancji:

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_s^2 & a\sigma_s^2 \\ a\sigma_s^2 & a^2\sigma_s^2 + \sigma_\delta^2 \end{bmatrix}.$$

Dla tak zdefiniowanego modelu do wyznaczenia estymatorów nieznanych parametrów można zastosować metodę największej wiarogodności, wprowadzając oznaczenia:

$$X_{i.} = \frac{1}{m}\sum_{j=1}^{m} X_{ij} \ ,$$

$$Y_{i.} = \frac{1}{m}\sum_{j=1}^{m} Y_{ij} \ ,$$

$$X_{..} = \frac{1}{n}\sum_{j=1}^{m} X_{i.} \ ,$$

$$Y_{..} = \frac{1}{n}\sum_{j=1}^{n} Y_{i.} \ ;$$

oraz

$$w_{xx} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - X_{i.})^2 \ ,$$

$$w_{yy} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - Y_{i.})^2 \ ,$$

$$w_{xy} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{ij} - X_{i.})(Y_{ij} - Y_{i.}) \ ,$$

$$s_{xx} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{i.} - X_{..})^2 \ ,$$

$$s_{yy} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{i.} - Y_{..})^2 \ ,$$

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}(X_{i.} - X_{..})(Y_{i.} - Y_{..}) \ ,$$

można wykazać, że estymatory największej wiarogodności nieznanych współczynników w modelu mają postać:

$$\hat{a} = \frac{w_{yy}s_{xx} - w_{xx}s_{yy} - \sqrt{\Delta}}{2(w_{xy}s_{xx} - w_{xx}s_{xy})} \ ,$$

$$\hat{b} = Y_{..} - \hat{a}X_{..} \ ,$$

gdzie: $\Delta = (w_{xx}s_{yy} - s_{xx}w_{yy})^2 - 4(s_{xy}w_{yy} - s_{yy}w_{xy})(s_{xx}w_{xy} - s_{xy}w_{xx})$.

Do wyrażenia formy estymatorów pozostałych parametrów wprowadzono oznaczenia:

$$B(\hat{a}) = s_{yy} - 2\hat{a}s_{xy} + \hat{a}^2 s_{xx},$$

$$W(\hat{a}) = w_{yy} - 2\hat{a}w_{xy} + \hat{a}^2 w_{xx},$$

oraz

$$p = (\hat{a}w_{xx} - w_{xy}) / W(\hat{a}),$$

$$q = (w_{yy} - \hat{a}w_{xy}) / W(\hat{a}).$$

Estymatory nieznanych wariancji zaburzeń zmiennej zależnej i zmiennej niezależnej mają odpowiednio postać:

$$\hat{\sigma}_\varepsilon^2 = p(W(\hat{a}) + B(\hat{a})) / \hat{a},$$

$$\hat{\sigma}_\delta^2 = q(W(\hat{a}) + B(\hat{a})).$$

Ponadto można wykazać, że dla każdego n, gdy $m \to \infty$ zastosowane w modelu estymatory metody największej wiarygodności są mocno zgodne, to oznacza, że dla każdego n wektor

$$\sqrt{m}\left[(a-\hat{a}),(b-\hat{b}),(\hat{\sigma}_\varepsilon - \sigma_\varepsilon),(\hat{\sigma}_\delta - \sigma_\delta),(\hat{\sigma}_s - \sigma_s),(\hat{s}_1 - s_1)\cdots,(\hat{s}_n - s_n)\right],$$

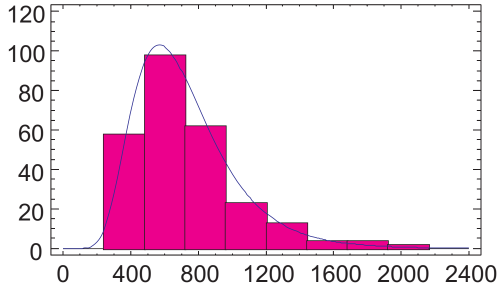
ma asymptotycznie rozkład normalny z zerowym wektorem wartości oczekiwanych.

3. Rozkład zmiennych losowych występujących w badaniu

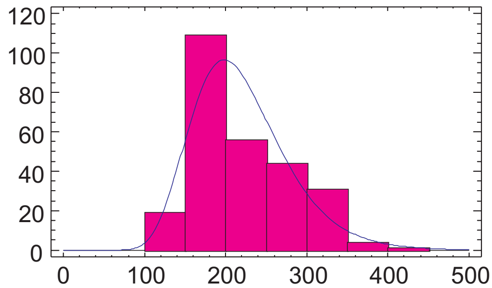
Przed przystąpieniem do konstrukcji modelu przeprowadzono badanie rozkładu zmiennych występujących w badanym modelu. Wartości dochodów i wydatków przedstawiono w cenach stałych, po uwzględnieniu inflacji. Dane są usystematyzowane ze względu na rodzaj gospodarstwa oraz na liczbę osób w gospodarstwie.

Analizie poddano gospodarstwa pracowników robotniczych i nierobotniczych, rolników, pracowników pracujących na własny rachunek oraz emerytów i rencistów. Gospodarstwa są podzielone na 1-, 2-, 3-, 4-, 5- oraz 6- i więcej osobowe. Niech D i W oznaczają odpowiednio zmienną określającą dochód wyrażony w cenach stałych oraz zmienną określającą wydatki na napoje alkoholowe, również wyrażone w cenach stałych. Ze względu na charakter opracowywania danych zamieszczonych w biuletynach, rozkład dopasowano do danych ważonych, w których wagami są liczby ankietowanych gospodarstw o danym charakterze.

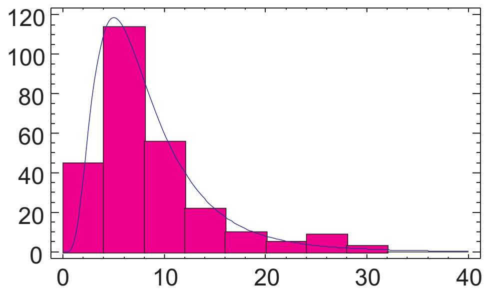
Na rysunkach 1–3 przedstawiono histogramy zmiennych oznaczających odpowiednio dochód *D* i wydatki *W* wraz z naniesionym oczekiwanym rozkładem logarytmiczno normalnym.



Rys. 1. Dopasowanie rozkładu logarytmiczno normalnego do zmiennej określającej wielkość dochodu na osobę



Rys. 2. Dopasowanie rozkładu logarytmiczno normalnego do zmiennej określającej wydatki na żywność i napoje bezalkoholowe na osobę



Rys. 3. Dopasowanie rozkładu logarytmiczno normalnego do zmiennej określającej wydatki na alkohol na osobę

4. Zależność między dochodem a wydatkami

W badaniach zależności pomiędzy dochodem a wydatkiem na pewne dobro konsumpcyjne zastosujemy dwa podejścia:
1) model z replikacjami,
2) klasyczny model prostej regresji liniowej.

Opierając się na analizie rozkładu zmiennych, zaproponowano model:

$$D_{ij} = P_{ij} \cdot e^{\varepsilon_{ij}} ,$$

$$W_{ij} = A \cdot P_{ij}^{B} \cdot e^{\sigma_{ij}} ,$$

w którym wskaźnik *i* przebiega struktury gospodarstw, a powtórzenia dokonywane są według lat.

Wielkości *A* i *B* są nieznanymi parametrami, natomiast zmienna P_{ij} określa prawdziwy nieznany średni dochód w danej grupie społecznej. Przyjęto, że rozkład P_{ij} jest logarytmicznonormalny, czyli

$$\ln(P_{ij}) = \xi_{ij} ,$$

gdzie: $\xi_{ij} \sim N\left(s_i, \sigma_s^2\right)$.

Ponadto:

$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$

oraz

$\delta_{ij} \sim N\left(0, \sigma_\delta^2\right)$.

Zakładamy, że wszystkie zmienne losowe są niezależne. Po przekształceniu otrzymujemy model, którego własności opisano w poprzednim paragrafie:

$$X_{ij} = \xi_{ij} + \varepsilon_{ij} ,$$

$$Y_{ij} = a\xi_{ij} + b + \delta_{ij} ,$$

gdzie:

$X_{ij} = \ln(D_{ij})$,
$Y_{ij} = \ln(W_{ij})$,
$\xi_{ij} = \ln(P_{ij})$.

Wówczas parametr *a* będzie zdefiniowanym wcześniej parametrem *B*, natomiast parametr *b* będzie oznaczać ln(*B*).

Badanie przeprowadzono w odniesieniu do wydatków na żywność i napoje bezalkoholowe.

4.1. Zależność między dochodem a wydatkami na żywność i napoje bezalkoholowe

Ze względu na charakter prezentowanych danych, które są podane z wagami uwarunkowanymi przez ilość gospodarstw o danym przekroju, w celu wyznaczenia estymatorów z modelu z replikacjami zastosowano estymatory ważone:

Przeprowadzając obliczenia za pomocą wzorów z poprzedniego paragrafu, otrzymano następujące wyniki, które zamieszczono w tabeli 1, co ostatecznie daje następującą zależność:

$$D_{ij} = P_{ij} \cdot e^{\varepsilon_{ij}},$$

$$W_{ij} = 3{,}311 \cdot P_{ij}^{0{,}634} \cdot e^{\delta_{ij}}.$$

Tabela 1

Obliczenia parametrów w ogólnym modelu z replikacjami

B = 0,634	Var(B) = 0,022	σ_ε = 0,124
ln(A) = 1,197	Var(ln(A)) = 0,142	σ_δ = 0,111

Jeśli rozważymy sytuację, w której zmienna niezależna (dochód) będzie podawana bez zaburzenia, wówczas do estymacji parametrów nieznanych parametrów zależności można zastosować metodę najmniejszych kwadratów. Omawianą zależność przedstawiono zatem następująco:

$$W_i = A \cdot D_i^B \, e^{\delta_i}.$$

Wyniki zamieszczono w tabeli 2.

Tabela 2

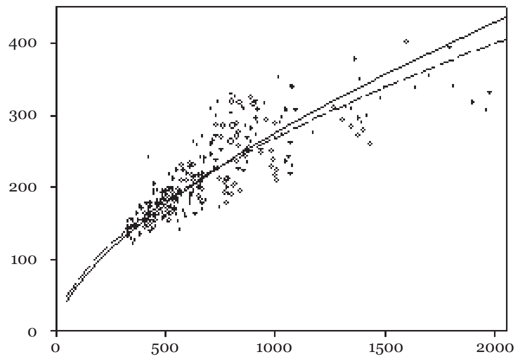
Obliczenia parametrów w modelu prostej regresji liniowej

B = 0,576	Var(B) = 0,021	
ln(A) = 1,613	Var(ln(A)) = 0,137	σ_δ = 0,108

Wnioskujemy zatem, że:

$$W_i = 4{,}211 \cdot D_i^{0{,}605} \, e^{\delta_i}.$$

Rysunek 4 przedstawia zależność wydatków na żywność i napoje bezalkoholowe od dochodu. Linia ciągła przedstawia zależność funkcyjną uzyskaną za pomocą modelu, w którym dochód jest zmienną losową (model z replikacjami), natomiast linia przerywana oznacza zależność funkcyjną, wynikającą z modelu prostej regresji.



Rys. 4. Zależność funkcyjna wydatków na żywność i napoje bezalkoholowe od dochodu

4.2. Zależność wydatków na alkohol od dochodu

Stosując podobną analizę w odniesieniu do wydatków na alkohol, otrzymujemy takie wyniki, jak te, które zestawiono w tabeli 3.

Tabela 3

Obliczenia parametrów w ogólnym modelu z replikacjami

B = 1,468	Var(B) = 0,042	σ_ε = 0,111
ln(A) = −7,584	Var(ln(A)) = 0,271	σ_δ = 0,203

Ostatecznie wprowadzamy zależność:

$$D_{ij} = P_{ij} \cdot e^{\varepsilon_{ij}},$$

$$W_{ij} = 5{,}08 \cdot 10^{-4} \cdot P_{ij}^{1{,}468} \cdot e^{\delta_{ij}}.$$

Jeśli rozważymy sytuację, w której zmienna niezależna, jaką jest dochód, jest bez zaburzenia, wówczas po zastosowaniu metody najmniejszych kwadratów otrzymano wyniki zestawione w tabeli 4.

Tabela 4

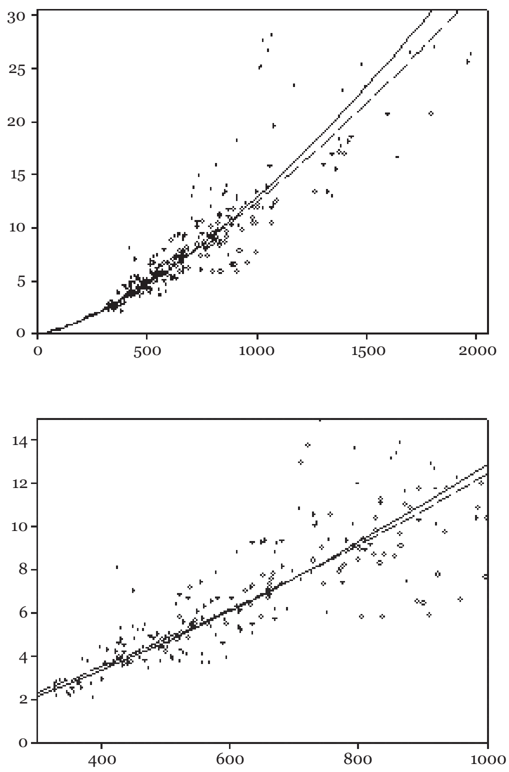
Obliczenia parametrów w modelu prostej regresji liniowej

V = 1,378	Var(B) = 0,037	
ln(A)= – 6,999	Var(ln(A)) = 0,238	σ_δ = 0,236

Na podstawie tych wyników:

$$W_i = 9{,}13 \cdot 10^{-4} \cdot W_i^{1{,}378}\, e^{\delta_i}.$$

Rysunek 5 przedstawia zależność wydatków na alkohol od dochodu. Linia ciągła przedstawia zależność funkcyjną uzyskaną stosując model, w którym dochód jest zmienną losową (model z replikacjami) natomiast linia przerywana oznacza zależność funkcyjną wynikającą z modelu prostej regresji.



Rys. 5. Zależność funkcyjna wydatków na alkohol od dochodu

5. Wnioski

Zbadano dwa przypadki. Analizując model dotyczący wydatków na żywność i napoje bezalkoholowe można zauważyć, że występuje tendencja do nasycenia, co oznacza, że zwiększanie dochodów powoduje relatywnie mały przyrost wydatków. Natomiast analizując drugi model, zauważamy, że zwiększanie dochodów skutkuje bardzo dużym przyrostem wydatków.

W obu przypadkach zastosowanie dwóch podejść do badania zależności dało różne wyniki. Dopuszczenie zaburzenia na zmiennej zależnej (dochód) pozwala zauważyć, że osoby o wyższym dochodzie wydają relatywnie więcej na żywność i napoje bezalkoholowe niż wynika to z analizy modelu, w którym nie zakładamy losowości zmiennej określającej dochód. Natomiast w przypadku osób z małym dochodem – odwrotnie. Osoby z niskimi dochodami wydają mniej niż to wynika z modelu regresji liniowej.

Literatura

[1] Bunke O., Bunke H., *Non-Linear Regression, Functional Relationships and Robust Methods*, Wiley, New York 1989.
[2] Chow C.G., *Ekonometria*, PWN, Warszawa 1995.
[3] Cox N.R., *The linear structural relation for several groups of data*, „Biometrica" 1976, No. 63, s. 231–237.
[4] Dolby G.R., *The ultrastructural relation a synthesis of the functional and structural relations*, „Biometrica" 1976, No. 63, s. 39–50.
[5] Friedman M., *A Theory of Consumption Function*, Princeton University Press, Princeton 1957.
[6] Fuller W.A., *Measurement Error Models*, Wiley, New York 1987.
[7] Piszczała J., *Matematyka i jej zastosowanie w naukach ekonomicznych*, Wyd. Akademii Ekonomicznej w Poznaniu, Poznań 1998.
[8] GUS, Budżety Gospodarstw Domowych, Warszawa 1996.