

PL ISSN 0071-674X

POLSKA AKADEMIA NAUK — ODDZIAŁ W KRAKOWIE
KOMISJA NAUK EKONOMICZNYCH I STATYSTYKI
KRAKOWSKA AKADEMIA
IM. ANDRZEJA FRYCZA MODRZEWSKIEGO

folia oeconomica cracoviensia

Vol. XLIX–L

2008–2009

WYDAWNICTWO ODDZIAŁU POLSKIEJ AKADEMII NAUK
KRAKÓW

POLSKA AKADEMIA NAUK — ODDZIAŁ W KRAKOWIE
KOMISJA NAUK EKONOMICZNYCH
KRAKOWSKA AKADEMIA
IM. ANDRZEJA FRYCZA MODRZEWSKIEGO

FOLIA OECONOMICA CRACOVIENSIA

Vol. XLIX–L
2008–2009

WYDAWNICTWO ODDZIAŁU POLSKIEJ AKADEMII NAUK
KRAKÓW

REDAKTOR

prof. dr hab. Andrzej Iwasiewicz

KOMITET REDAKCYJNY

prof. dr hab. Anna Czubała

prof. dr hab. Henryk Gurgul

prof. dr hab. Jacek Osiewalski — Sekretarz Naukowy Komisji Nauk Ekonomicznych
i Statystyki Oddziału PAN w Krakowie i sekretarz naukowy Komitetu Redakcyjnego

Adres redakcji

31-018 Kraków, ul. św. Jana 28

Wydanie publikacji finansowane przez
Ministra Nauki i Szkolnictwa Wyższego,
Krakowską Akademię im. A. Frycza Modrzewskiego

Redaktor Wydawnictwa

Krystyna Duszyk

© Copyright by Autorzy, Polska Akademia Nauk Oddział w Krakowie
Kraków 2009

ISSN 0071-674X

Wydawnictwo Oddziału Polskiej Akademii Nauk
31-018 Kraków, ul. św. Jana 28
tel. 012 422-36-43 w. 12, 22, kolportaż 15; fax 012 422-27-91
www.pan-krakow.pl
Ark. wyd. 12; ark. druk. 10,5
Druk: PANDIT
31-334 Kraków, ul. Łokietka 177



SPIS TREŚCI

Daniel Kosiorowski: <i>Wybrane zagadnienia koncepcji głębi danych</i>	5
Henryk Gurgul, Tomasz Wójtowicz: <i>Stopy zwrotu a wielkość obrotów na GPW w Warszawie ...</i>	31
Jacek Osiewalski, Renata Wróbel-Rotter: <i>Bayesowskie graniczne funkcje kosztu dla sektora dystrybucji energii</i>	47
Andrzej Iwasiewicz: <i>Monitorowanie procesów binarnych za pomocą kart kontrolnych sum skumulowanych</i>	71
Tadeusz Grabiński: <i>Uwarunkowania pomiaru współzależności cech</i>	91
Łukasz Kwiatkowski: <i>Markov switching in stochastic variance. Bayesian comparison of two simple models</i>	109
Jacek Kwiatkowski: <i>Model Stocka i Watsona oraz jego modyfikacje — analiza inflacji w Polsce</i>	145

WYBRANE ZAGADNIENIA KONCEPCJI GŁĘBI DANYCH

DANIEL KOSIOROWSKI

Katedra Statystyki
Uniwersytet Ekonomiczny w Krakowie
PL 31-510 Kraków, ul. Rakowicka 27
e-mail: daniel.kosiorowski@uek.krakow.pl

Praca przedstawiona przez Autora na posiedzeniu Komisji Nauk Ekonomicznych Oddziału PAN w Krakowie w dniu 4 grudnia 2007 r.

ABSTRACT

D. Kosiorowski. *Selected Issues of Data Depth Concept*. Folia Oeconomica Cracoviensia 2008-2009, 49-50: 5-30.

In this paper we present selected aspects of data depth concept. We propose several statistical procedures based on data depth concept. We study a performance of the propositions on various multivariate data sets simulated from skewed, fat tailed distributions and mixtures of them.

KEY WORDS — SŁOWA KLUCZOWE

robust statistical procedure, statistical depth function
odporną procedurą statystyczną, statystyczna funkcja głębi

1. WPROWADZENIE

Koncepcja głębi danych jest jednym z trzech dominujących nurtów badań nad pojęciem i zastosowaniami wielowymiarowego kwantyla (patrz np. Liu i in., 1999).

Wykazuje szereg związków z tzw. koncepcją głębi regresyjnej (ang.: *regression depth*) (np. Mizera, 2002) oraz staje się coraz atrakcyjniejszą alternatywą dla historycznie pierwszej koncepcji kwantyli przestrzennych (ang. *spatial (geometrical) quantile*) (np. Chaudhuri, 1996). Kwantyle definiowane w obrębie trzech podejść różnią się interpretacją.

Narzędzia oferowane przez koncepcję głębi danych wykorzystywane są w zagadnieniach nieparametrycznego wnioskowania statystycznego wykorzystującego pojęcia rangi, statystyki porządkowej, kwantyla itp. Dodajmy, że w obrębie tzw. klasycznego podejścia do nieparametrycznego wnioskowania w R^d , $d > 1$, wykorzystuje się wektory jednowymiarowych statystyk. Postępowanie takie nie uwzględnia często szczególnie istotnej geometrii wielowymiarowego zbioru danych. Na przykład, może się zdarzyć, że wektor jednowymiarowych median leży poza powłoką wypukłą wielowymiarowego zbioru danych — nie może zatem być dobrą miarą położenia centrum.

Charakterystyczną dla koncepcji miarę centralności punktu $x \in R^d$, $d > 1$, będącego realizacją pewnego d -wymiarowego wektora losowego X o rozkładzie prawdopodobieństwa P , wprowadza się za pomocą specjalnej funkcji nazywanej *głębiami* (ang. *depth*) bądź funkcją *głębii* (ang. *depth function*).

Funkcja *głębii* przyporządkowuje każdemu punktowi liczbę rzeczywistą z przedziału $[0, 1]$, będącą miarą jego centralności, zważywszy na rozkład go generujący.

Wykorzystując funkcję *głębii* można uporządkować zbiór wielowymiarowych obserwacji na zasadzie odstawiania obserwacji od centrum.

Zazwyczaj punkt, dla którego funkcja *głębii* przyjmuje wartość maksymalną określa się mianem *d -wymiarowej mediany*.

Oznaczamy przez P rodzinę rozkładów prawdopodobieństwa określonych na σ ciele zbiorów borelowskich w R^d oraz przez P_X rozkład danego wektora losowego X .

Każdy element próby $X^n = \{X_1, \dots, X_n\}$ traktujemy jako $d \times 1$ wektor kolumnowy. Zakładamy ponadto, że rozkład P_X jest absolutnie ciągły.

W literaturze przedstawiono kilka układów postulatów, które powinna spełniać funkcja *głębii*, aby była odpowiednim narzędziem służącym do budowy nieparametrycznych wielowymiarowych procedur statystycznych (zob. Dycerhoff, 2004; Zuo, i Serfling, 2000. Poniżej przedstawiamy najczęściej stosowany układ postulatów, który spełnia większość znanych funkcji *głębii* $D(x, P)$.

1. Niezmienniczość afiniczna — $D(x, P)$ jest niezależna od wyboru układu współrzędnych.

2. Wartość maksymalna w centrum — jeżeli rozkład P jest symetryczny względem Θ w pewnym sensie, wówczas $D(x, P)$ przyjmuje w tym punkcie maksimum.

3. Symetria — jeżeli rozkład P jest symetryczny względem Θ w pewnym sensie, wtedy także $D(x, P)$ jest symetryczna w tym sensie.

4. Zmniejszanie się wartości wzdłuż promieni — wartość funkcji *głębii* $D(x, P)$ zmniejsza się wzdłuż promienia mającego początek w punkcie o maksymalnej *głębii*.

5. Zanikanie w nieskończoności — $D(x, P) \rightarrow 0$ gdy $\|x\| \rightarrow \infty$.

6. Ciągłość $D(x, P)$ jako funkcji x .

7. Ciągłość $D(\mathbf{x}, P)$ rozpatrywanej jako funkcjonal P .

8. Quasi-wypukłość $D(\mathbf{x}, P)$ rozpatrywanej jako funkcja \mathbf{x} — zbiór $\{\mathbf{x} : D(\mathbf{x}, P) \geq \alpha\}$ jest wypukły dla każdego $\alpha \in [0,1]$.

Zaznaczmy, że przez centrum rozumiemy punkt symetrii, przez symetrię rozumiemy centralną symetrię.

Zbiór

$$\{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}, P) = \alpha\}, \quad (1)$$

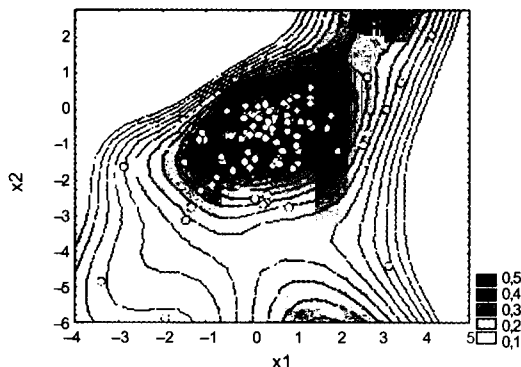
nazywany poziomem bądź konturem głębzi α , zbiór ten określany jest mianem d -wymiarowego kwantyla rzędu α , $\alpha \in [0, 1]$.

Zbiór

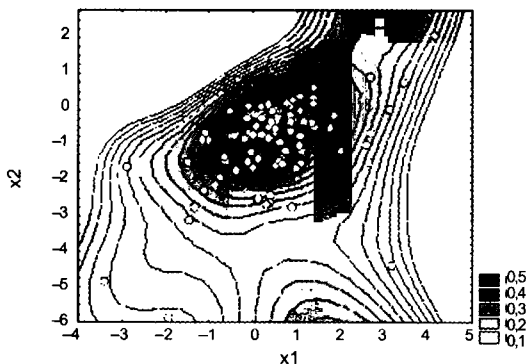
$$D_\alpha(X) = \{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}, P) \geq \alpha\}, \quad (2)$$

tzn. obszar ograniczony przez kontur głębzi α , nazywany jest α przyciętym (centralnym) obszarem $\alpha \in [0,1]$.

Opierając się na danej funkcji głębzi, możemy zdefiniować punkt o maksymalnej głębzi jako wielowymiarowy analog jednowymiarowej mediany.



Ryc. 1. Projekcyjne obszary centralne ze 100 elementowej próby skośnego rozkładu normalnego z $\Omega = \text{diag}(2) \cdot 5$, $\mathbf{m} = (0, 0)$, $\alpha = (2, -5)$



Ryc. 2. Projekcyjne obszary centralne ze 100 elementowej próby z rozkładu Marshalla-Olkina $\lambda = (1, 1, 1)$

Niech będzie funkcją głębi, wtedy medianę indukowaną przez definiujemy:

$$M(P) = \operatorname{argsup}_{x \in \mathbb{R}^d} D(x, P). \quad (3)$$

Przykłady statystycznych funkcji głębi

Symetryczna głębia projekcyjna punktu $x \in \mathbb{R}^d$, będącego realizacją pewnego d -wymiarowego wektora losowego X o rozkładzie prawdopodobieństwa F , $PD(x, F)$ definiowana jest jako:

$$PD(x, F) = \left(1 + \sup_{\|u\|=1} \frac{|Med(u^t X)|}{MAD(u^t X)} \right)^{-1}, \quad (4)$$

gdzie X ma rozkład prawdopodobieństwa F , Med oznacza jednowymiarową medianę, oraz MAD oznacza jednowymiarową medianę odchylenia absolutnego od mediany $MAD(Z) = Med(|Z - Med(Z)|)$.

Głębia projekcyjna oraz indukowane przez nią estymatory położenia centrum oraz rozrzutu wektora losowego odznaczają się bardzo dobrymi własnościami w kategoriach odporności oraz efektywności dla szerokiej klasy populacji. Głębia ta jest afinicznie niezmiennicza.

Głębia Tukey'a (głębia domkniętej półprzestrzeni) definiowana jest:

$$D(x, P) = \inf_H \{P(H) : x \in H, H \text{ jest domkniętą półprzestrzenią}\}. \quad (5)$$

Głębia punktu x — najmniejsza masa probabilistyczna znajdująca się na domkniętej półprzestrzeni z punktem x na brzegu.

Najprostszym przykładem głębi jest tzw. głębia Euklidesa:

$$D_{EUK}(z | X^n) = \frac{1}{1 + \|z - \bar{x}\|^2}, \quad \text{gdzie } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (6)$$

Znaną głębię stanowi tzw. głębia Mahalanobisa:

$$D_{MAH}(y | X^n) = \frac{1}{1 + (y - \bar{y})^t S^{-1} (y - \bar{y})}, \quad (7)$$

gdzie S oznacza macierz kowariancji próby X^n .

Pojęcie głębi dopasowania funkcji regresji liniowej zostało wprowadzone przez Rousseeuw i Hubert (Rousseeuw i Hubert, 1998).

I. Mizera (2000) uogólnił ich podejście oraz pokazał je w ramach formalizmu optymalizacji wektorowej uzyskując jednocześnie szereg wyników dotyczących odporności estymatorów maksymalnej głębi regresyjnej.

Przypuśćmy, że zamierzamy dopasować liniową funkcję regresji $y = b_1 x + b_2$ do dwuwymiarowego zbioru danych $Z^n = \{(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)\}$.

Oznaczmy współczynniki regresji jako $\mathbf{b} = (b_1, b_2)$. Oznaczmy reszty regresji jako:

$$r_i(\mathbf{b}) = r_i = y_i - b_1 x_1 - b_2.$$

DEFINICJA 1. Powiemy, że dopasowanie $\mathbf{b} = (b_1, b_2)$ nie jest słabo optymalne względem zbioru danych \mathbf{Z}^n jeżeli istnieje liczba rzeczywista $v_b = v$, która nie pokrywa się z żadnym punktem x_i i dla której zachodzi:

$$r_i(\mathbf{b}) < 0 \text{ dla wszystkich } x_i < v \text{ i } r_i(\mathbf{b}) > 0 \text{ dla wszystkich } x_i > v$$

lub:

$$r_i(\mathbf{b}) > 0 \text{ dla wszystkich } x_i < v \text{ i } r_i(\mathbf{b}) < 0 \text{ dla wszystkich } x_i > v$$

(istnienie liczby v odpowiada istnieniu punktu, dookoła którego możemy obracać prostą do pozycji pionowej, nie napotykając żadnej obserwacji).

DEFINICJA 2. Głębina regresyjna $rdepth(\mathbf{b}, \mathbf{Z}^n)$ jest najmniejszą frakcją obserwacji próby \mathbf{Z}^n , którą należy usunąć aby dopasowanie \mathbf{b} przestało być słabo optymalne.

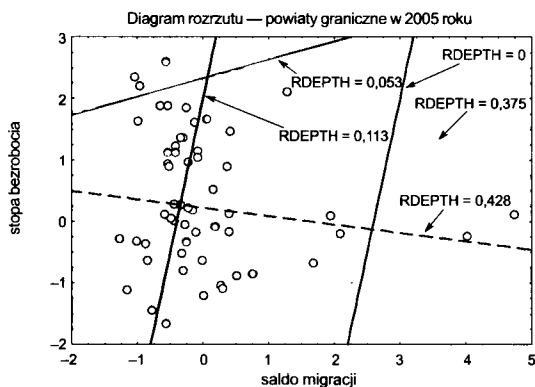
Estymator maksymalnej głębiny regresyjnej definiujemy jako:

$$T_r^*(\mathbf{Z}^n) = \arg \max_{\mathbf{b}} rdepth(\mathbf{b}, \mathbf{Z}^n) = \arg \max_{\mathbf{b}^{ij}} rdepth(\mathbf{b}^{ij}, \mathbf{Z}^n). \quad (8)$$

Warto zauważyć, że w przypadku koncepcji głębiny regresyjnej nie czyni się jakichkolwiek założeń odnośnie regularności składnika losowego, wpływającego na wartości obserwacji.

Dopasowanie największej głębiny jest prostą najlepiej równoważącą chmurę danych.

Zaproponowane przez Mizere ogólniejsze spojrzenie na koncepcję głębiny regresyjnej z jednej strony pokazuje jej miejsce w ramach całej koncepcji głębiny danych, z drugiej strony — w wielu przypadkach prowadzi do prostszych metod badania własności estymatorów maksymalnej głębiny.



Ryc. 3. Stopa bezrobocia vs. saldo migracji w polskich powiatach granicznych w 2005 roku. Głębina dopasowań funkcji regresji do danych empirycznych. Linia przerywana reprezentuje estymator maksymalnej głębiny, linia oznaczona kropkami reprezentuje estymator metody heteroskedastycznej regresji *t*-Studenta

2. KRZYWA SKALI — ODPORNA I NIEPARAMETRYCZNA METODA BADANIA ROZRZUTU WEKTORA LOSOWEGO I STOPNIA ZALEŻNOŚCI ROZKŁADÓW BRZEGOWYCH

Wskazanie odpornej i zarazem efektywnej alternatywy dla wektora przeciętnych oraz macierzy kowariancji jako estymatorów, odpowiednio, położenia centrum i rozrzutu wektora losowego należy do najważniejszych celów współczesnej wielowymiarowej analizy statystycznej. Macierz kowariancji z próby ma nieograniczona funkcję wpływu Hampela, co znaczy, że nie jest odporna na lokalne punktowe zmieszania. Punkt załamania (BP) próby skończonej Donoho i Hubera macierzy kowariancji z n -elementowej próby wynosi $1/n$, zaledwie jedna obserwacja odstająca jest w stanie istotnie zniekształcić ocenę rozrzutu rozpatrywanego wektora. Macierz kowariancji z próby ma nieograniczone maksymalne obciążenie Hubera, czyli nie jest odporna m.in. na błędną specyfikację modelu generującego obserwacje. Praktyczne wykorzystanie macierzy kowariancji z próby wiąże się z istnieniem momentów drugiego stopnia wektora losowego, reprezentującego badane zjawisko. Interpretacja macierzy kowariancji jest utrudniona w przypadku skośnych populacji.

Wykorzystując głębię projekcyjną $PD(x, F)$ definiuje się tzw. projekcyjne obszary centralne rzędu r (w obrębie koncepcji głębi danych ich brzegi określa się mianem d -wymiarowych kwantyli):

$$PC_F(r) = \{x : PD(x, F) \geq r\}. \quad (9)$$

W przypadku, gdy rozkład F jest centralnie symetryczny, obszary centralne odznaczają się taką samą własnością.

Wykorzystując obszary projekcyjne centralne możemy zdefiniować tzw. krzywą skali, będącą rzeczywistym funkcjonałem objętości obszarów centralnych, a służącą do nieparametrycznego opisu rozrzutu wektora losowego wokół wielowymiarowej mediany.

Krzywa skali definiowana jest jako:

$$v_F(r) = \Delta(PC_F(r)), \quad 0 \leq r < 1, \quad (10)$$

gdzie: $\Delta(\cdot)$ oznacza miarę Lebesgue'a a $PC_F(\cdot)$ — projekcyjny obszar centralny.

Krzywa skali jest dwuwymiarową metodą opisu rozrzutu wartości wektora losowego wokół mediany projekcyjnej. W związku z faktem, że projekcyjne obszary centralne stanowią zagnieżdżoną rodzinę zbiorów, krzywa skali służy do pomiaru stopnia ekspansji obszarów centralnych wraz ze wzrastającą masą probabilistyczną w nich zawartą.

Niech F_0 będzie „rozkładem niezależności”¹ związanym z danym rozkładem F . Łatwo zauważyć, że krzywa skali F_0 powinna przebiegać powyżej krzy-

¹ Polski odpowiednik niewątpliwie wymaga dopracowania.

wej skali F . Można wykorzystać obszar pomiędzy krzywą skali F_0 i krzywą skali F do pomiaru stopnia zależności rozkładów brzegowych F .

M. Romanazzi (2004) sugeruje, aby w tym celu wykorzystać znormalizowaną wersję odległości Euklidesa pomiędzy krzywymi skali nazywaną krzywą korelacji:

$$\gamma_1(\alpha, F) = \left(\frac{\int_0^\alpha (\Delta C(t, F_0) - \Delta C(t, F))^2 dt}{\int_0^\alpha (\Delta C(t, F_0) - \Delta C(t, F))^2 dt} \right)^{1/2}, \quad (11)$$

gdzie:

$\Delta(\cdot)$ oznacza miarę Lebesgue'a, F_0 jest „rozkładem niezależności”.

Krzywa korelacji wyraża odległość F od rozkładu niezależności F_0 dla $0 \leq \alpha < 1$, aby przedstawić ją graficznie sporządzamy diagram rozrzutu $\gamma_i(\alpha, F)$ vs. α .

Najczęściej nie jesteśmy w stanie wskazać „rozkładu niezależności”, gdyż nie znamy klasy rozkładów, do której należy rozkład generujący dane. Okazuje się, że nawet w takich sytuacjach można z powodzeniem wykorzystywać krzywą korelacji.

Przypuśćmy, że X jest $n \times p$ macierzą losową, której wiersze $X_i = (X_{i1}, \dots, X_{ip})$ są obserwacjami z n -elementowej próby losowej z p -wymiarowego rozkładu F oraz rozważmy odwzorowanie $X \rightarrow \tau X$, gdzie τ jest przekształceniem, które zamienia każdą kolumnę $X^{(j)} = (X_{1j}, \dots, X_{nj})^T$ macierzy X permutacją jej składowych. Zauważmy, że jeśli kolumny mają różne elementy, wtedy jest takich przekształceń.

Niech \hat{F}_m będzie rozkładem τX oraz niech X będzie klasą takich rozkładów. Romanazzi (2004) dowodzi twierdzenia głoszącego, że warunkując obserwowaną próbą stosownym „rozkładem niezależności” dla F jest mieszanina:

$$\hat{F}_{n,0} = \frac{1}{(n!)^p} \sum_{\tau \in C} \hat{F}_{n,\tau}. \quad (12)$$

Aproksymację $\tilde{F}_{n,0}$ uzyskuje się biorąc losową próbę m rozkładów z X .

Można pokazać, że dla $F \sim N_p(\mathbf{m}, \Sigma)$, gdzie Σ jest dodatnio określoną macierzą wymiaru $p \times p$ dla $0 < \alpha < 1$ zachodzi:

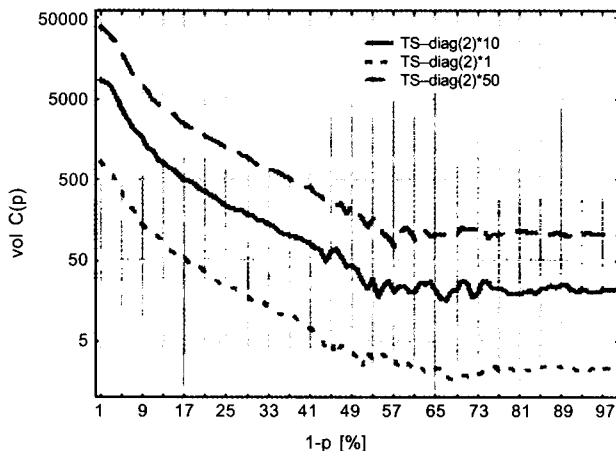
$$\gamma_1(\alpha, F) = (1 - (\det R)^{1/2}) / (1 + (\det R)^{1/2}), \quad (13)$$

gdzie $R = (\text{diag} \Sigma)^{-1/2} \Sigma (\text{diag} \Sigma)^{-1/2}$ jest macierzą korelacji F .

W celu sprawdzenia wybranych statystycznych własności krzywych skali i korelacji z próby przeprowadzono badania symulacyjne. Generowano mianowicie po 500 prób złożonych ze 100 obserwacji pochodzących z dwuwymiarowych rozkładów skośnych normalnych, skośnych T -Studenta, Marshalla-Olkina oraz mieszanin tychże rozkładów. Eksperymenty powtarzano kilkadziesiąt razy dla sprawdzenia stabilności oszacowań.

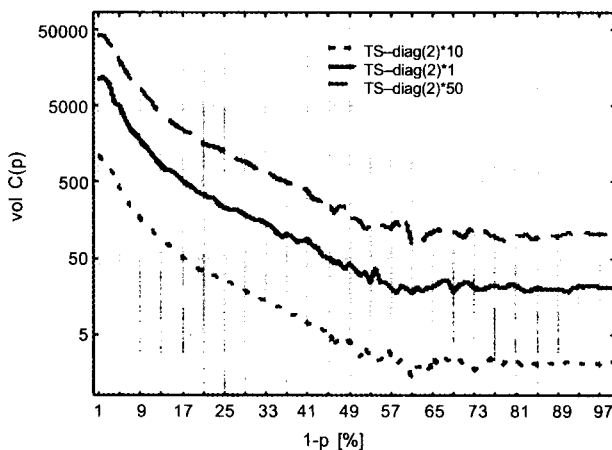
Uzyskane wyniki skłaniają do następujących wniosków:

- Krzywe skali dobrze dyskryminują rozkłady skośne normalne i skośne T -Studenta różniące się charakterystykami rozrzutu.
- Krzywe korelacji sporządzone dla izotropowych skośnych rozkładów normalnych i T właściwie „wychwytyją” wpływ skośności na brak niezależności rozkładów brzegowych.
- Krzywa korelacji dobrze się sprawuje w przypadku rozkładu nie należącego do rodziny wykładniczej. Należy podkreślić, że z krzywej korelacji można odczytać jak przedstawia się struktura zależności rozkładu w zależności od bliskości centrum rozkładu.



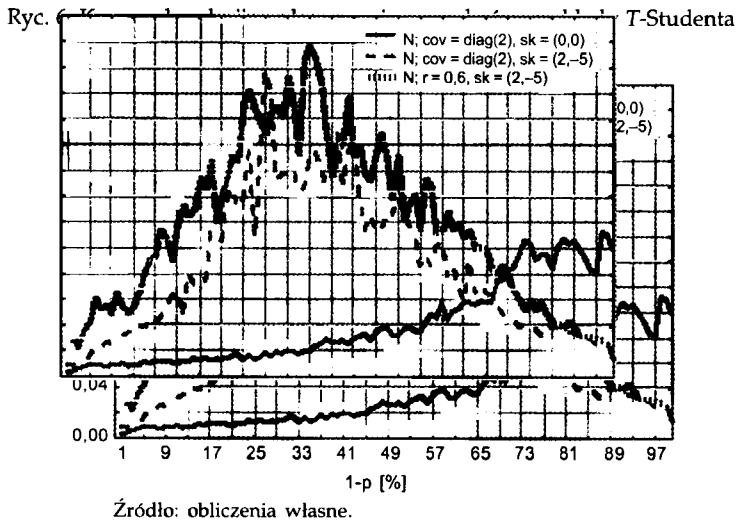
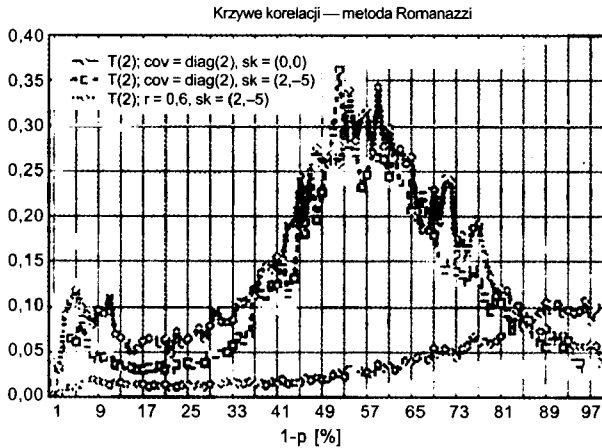
Źródło: obliczenia własne.

Ryc. 4. Krzywe skali — dwuwymiarowe rozkłady normalne



Źródło: obliczenia własne.

Ryc. 5. Krzywe skali — dwuwymiarowe rozkłady T -Studenta



Ryc. 7. Krzywe korelacji — dwuwymiarowe skośne rozkłady normalne

- Krzywa korelacji jest wrażliwa na zmieszanie populacji, co w zależności od punktu widzenia można poczytać za jej wadę bądź zaletę.
- Krzywa skali jest względnie niewrażliwa na zmieszania populacji.

3. KWANTYLOWY FUNKCJONAŁ ASYMETRII ROZKŁADU WEKTORA LOSOWEGO

Odstępstwo wielowymiarowego rozkładu prawdopodobieństwa od ustalonego pojęcia symetrii określa się mianem skośności rozkładu. Przez symetrię na ogół rozumie się własność obiektu polegającą na tym, że istnieje pewne,

różne od tożsamościowego przekształcenie, które odwzorowuje dany obiekt na niego samego.

Zaznaczmy, że w przypadku wielowymiarowego rozkładu prawdopodobieństwa wykorzystuje się wiele różniących się wzajemnie pojęć wielowymiarowej symetrii, które sprowadzają się do zwykłego pojęcia symetrii w przypadku jednowymiarowym tzn. symetrii zwierciadlanej.

Mówimy, że wektor losowy X ma rozkład centralnie symetryczny względem θ , jeżeli:

$$X - \theta \stackrel{r}{=} \theta - X.$$

gdzie symbol „ $\stackrel{r}{=}$ ” oznacza równość rozkładów.

Propozycja nawiązująca do koncepcji głębi

Stopień odstępstwa rozkładu prawdopodobieństwa od centralnej symetrii można mierzyć za pomocą funkcjonału asymetrii wykorzystującego stosownie wybraną funkcji głębi, np. głębię projekcyjnej.

Niech PM_F oznacza medianę indukowaną przez głębię projekcyjną. Aby zmierzyć asymetrię rozkładu dla każdego $r \in (0, 1)$, badamy różnicę pomiędzy przeciętną punktów wewnątrz obszaru centralnego rzędu r a medianą PM_F .

Rozważmy mianowicie:

$$\|\tilde{s}_F(p)\| = 2 \left\| \frac{\int_{PC_F(p)} W(x) m_p(dx) - PM_F}{\tilde{v}_F(p)^{1/d}} \right\|, \quad 0 < p < 1, \quad (14)$$

gdzie: $PC_F(p)$ to projekcyjny obszar centralny rzędu r , $m_p(dx)$ oznacza rozkład, np. jednostajny na $PC_F(p)$, PM_F oznacza indukowaną przez głębię projekcyjną medianę, $W(\cdot)$ właściwą dla zagadnienia funkcję wagową np. $W(x) = x$.

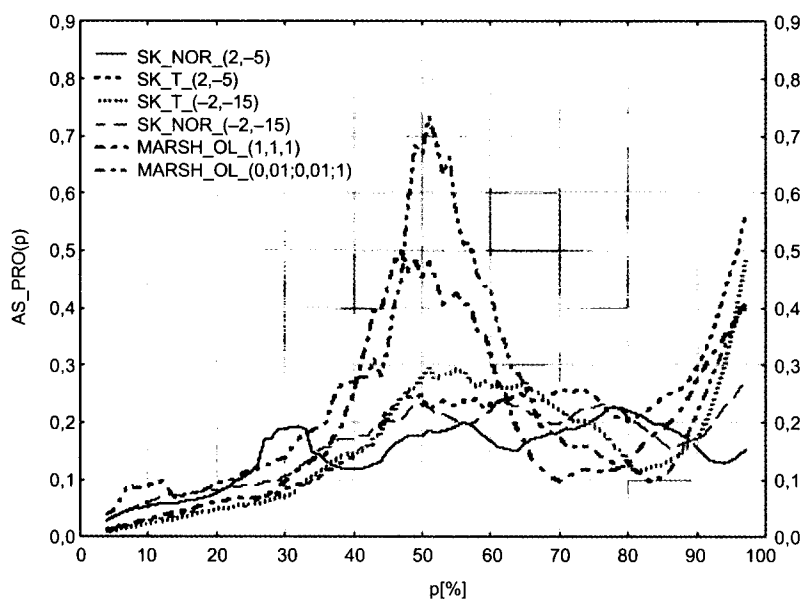
Proponowany funkcjonał asymetrii, dzięki własnościom głębi projekcyjnej i własnościom indukowanych przez nią obszarów centralnych, jest afinicznie niezmienniczy, tzn. nie zależy od przyjętego w badaniu układu współrzędnych. Umożliwia nieparametryczny pomiar asymetrii populacji nie posiadającej momentów. Można pokazać, że funkcjonał (14) z próby jest mocno zgodnym w sensie odległości Kołmogorowa estymatorem odpowiednika w populacji.

W celu sprawdzenia wybranych własności (14) przeprowadzono badania symulacyjne. Generowano mianowicie po 100 prób 100 elementowych z dwuwymiarowych rozkładów:

a) skośnego normalnego i skośnego T o dwóch stopniach swobody o parametrach: położenia $\mathbf{m} = (0, 0)$, rozrzutu $\Sigma = \text{diag}(2) \cdot 5$, kształtu $\Omega = (2, -5)$;

b) skośnego normalnego i skośnego T o dwóch stopniach swobody o parametrach: położenia $\mathbf{m} = (0, 0)$, rozrzutu $\Sigma = \text{diag}(2) \cdot 5$, kształtu $\Omega = (-2, -15)$;
 c) Marshalla-Olkina o parametrach $\lambda_1 = (1, 1, 1)$ i $\lambda_2 = (0,01, 0,01, 1)$.

Z ryciny 8 wynika m.in., że proponowany funkcjonal dobrze rozróżnia zarówno wyszczególnione typy rozkładów (skośny normalny, skośny T , Marshalla-Olkina), jak i rozkłady należące do tego samego typu a różniące się parametrem asymetrii. Dodajmy, że z prowadzonych wcześniej badań wynika, że nasza propozycja lepiej rozróżnia pomiędzy skośnymi rozkładami normalnym i T aniżeli oryginalna propozycja Chaudhuri (1996) oraz propozycje Liu i in (1999).



Ryc. 8. Wyniki badań symulacyjnych proponowanego funkcjonalu asymetrii

4. ODPORNOŚĆ METOD KLASYFIKACYJNYCH WYKORZYSTUJĄCYCH FUNKCJE GŁĘBI

Rozważamy k populacji p wymiarowych C_1, \dots, C_k , $k \geq 2$. Przypuśćmy, że z każdą populacją C_j związana jest gęstość prawdopodobieństwa $f_j(\mathbf{z})$ na R^p , w ten sposób, że jeśli obserwacja pochodzi z populacji C_j , to została wygenerowana przez rozkład o gęstości $f_j(\mathbf{z})$.

W tzw. zagadnieniu dyskryminacji obiektów interesuje nas racjonalny sposób przyporządkowywania obserwacji $\mathbf{z} \in R^p$ do jednej ze wspomnianych k populacji (zob. np. Krzyśko, 2006; Jajuga, 1993).

Reguła dyskryminacyjna L odpowiada podziałowi R^p na rozłączne obszary R_1, \dots, R_k , spełniające warunek $\bigcup R_j = R^p$. Reguła definiowana jest jako:

Przyporządkuj z do C_j jeżeli $z \in R_j$, dla $j = 1, \dots, k$.

Indeks $i \in \{1, 2, \dots, k\} = Y$ wskazujący na rozważaną populację C_i często określa się mianem etykiety populacji. W takim ujęciu zagadnienie dyskryminacji sprowadza się do predykcji etykiety $i \in Y$ na podstawie obserwacji wektora cech z .

Reguła klasyfikacyjna nazywana klasyfikatorem jest zatem funkcją:

$$L : R^p \ni x \longrightarrow i \in Y.$$

Gdy obserwujemy wektor $x \in X$, to prognozą jego przynależności jest $L(x) \in Y$.

Często w praktyce rozważa się sytuację, gdy wprowadzie ogólne postaci gęstości $f_j(z)$ są znane jednak jesteśmy zmuszeni szacować ich parametry. Estymacja w takim przypadku opiera się na tzw. próbie uczącej — macierzy danych $Z_{n \times p}$, której wiersze są podzielone na k grup odpowiadających k rozpatrywanym populacjom, macierz zostaje podzielona na $k(n_j \times p)$ macierzy Z_j odpowiada próbie n_j obserwacji z populacji C_j .

Klasyczne metody dyskryminacyjne, takie jak liniowe bądź kwadratowe funkcje dyskryminacyjne, zakładają wielowymiarową normalność (szerzej eliptryczność) populacji. Metody te nie sprawują się dobrze w przypadku skośnych populacji. Metody klasyczne zakładają, że rozpatrywane populacje posiadają momenty. Fakt ten stanowi istotne ograniczenie ich stosowalności, np. w przypadku wielowymiarowego rozkładu Cauchy'ego. Metody wykorzystujące wektory przeciętnych, macierze kowariancji bądź kryterium najmniejszych kwadratów są skrajnie nieodporne na jednostki odstające.

Odporność reguły dyskryminacyjnej

DEFINICJA 3 (na podst. Krzyśko, 2006): Rozważamy k populacji p wymiarowych C_1, \dots, C_k , $k \geq 2$ oraz ustaloną próbę uczącą reprezentującą populacje. Rzeczywisty poziom błędu klasyfikatora L jest równy:

$$Err(L) = P\{L(X) \neq i \mid X \in C_i\},$$

gdzie X oznacza obserwację niezależną od próby uczącej.

$Err(L)$ jest prawdopodobieństwem zdarzenia, że klasyfikator błędnie zakwalifikuje nową obserwację niezależną od próby uczącej pod warunkiem, że próba ucząca jest ustalona.

Propozycja. Rozważamy k populacji p wymiarowych C_1, \dots, C_k , $k \geq 2$, oraz próbę uczącą Z reprezentującą populacje. Punkt załamania próby uczącej Z klasyfikatora L w j -tej klasie C_j określamy jako:

$$BP_j(L, C_j^m) = \inf_{C_j^m} \left(\frac{m}{n_j} : \frac{P\{L(\mathbf{z}) \neq i \mid \mathbf{z} \in C_i\}}{P\{L(\mathbf{z}) = i \mid \mathbf{z} \in C_i\} + P\{L(\mathbf{z}) \neq i \mid \mathbf{z} \in C_i\}} > \frac{1}{2} \right), \quad (15)$$

gdzie C_j^m oznacza $(n_j \times p)$ macierz Z_j próby uczącej Z odpowiadającą próbie n_j obserwacji z populacji C_j , w której zastąpiono m wierszy (m obserwacji) dowolnymi wierszami (obserwacjami).

Globalny punkt załamania próby uczącej klasyfikatora określamy jako:

$$BP(L, C_1, \dots, C_k) = \min_j BP_j(L, C_j^m). \quad (16)$$

Każda funkcja głębi D indukuje pewną regułę klasyfikacyjną, tzn.:

$$L(\mathbf{z}) = \operatorname{argmax}_j D(\mathbf{z} \mid C_j), \quad j = 1, \dots, k, \quad (17)$$

która klasyfikuje obserwację do tej klasy C_j , dla której głębia z przyjmuje wartość maksymalną (zob. Hoberg i Mosler, 2006).

Przykłady reguł klasyfikacyjnych indukowanych przez głębie

Niech $C_j = \{\mathbf{x}_{j1}, \dots, \mathbf{x}_{jn_j}\}$ oznacza próbę n_j obserwacji z populacji j , $j = 1, \dots, k$.

Przykładem reguły klasyfikacyjnej indukowanej przez funkcję głębi jest reguła wykorzystująca tzw. głębię Euklidesa:

$$D_{EUK}(\mathbf{z} \mid C_j) = \frac{1}{1 + \|\mathbf{z} - \bar{\mathbf{x}}_j\|^2}, \quad \text{gdzie } \bar{\mathbf{x}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{x}_{ji}.$$

Zauważmy, że głębia ta prowadzi do znanej reguły k -centroidów.

Wykorzystując znaną głębię Mahalanobisa:

$$D_{MAH}(\mathbf{z} \mid C_j) = \frac{1}{1 + (\mathbf{y} - \bar{\mathbf{x}})' S_j^{-1} (\mathbf{y} - \bar{\mathbf{x}})},$$

gdzie S_j oznacza macierz kowariancji próby C_j , otrzymamy regułę klasyfikacyjną Mahalanobisa.

Propozycja. Rozważmy regułę klasyfikacyjną indukowaną przez tzw. symetryczną głębię projekcyjną (własności tej głębi przedstawia m.in. Zuo, 2003)

$$D_{PRO}(\mathbf{z} \mid C_j) = \left(1 + \sup_{\|\mathbf{u}\|=1} \frac{|\mathbf{u}'\mathbf{z} - \operatorname{med}(\mathbf{u}'C_j)|}{MAD(\mathbf{u}'C_j)} \right)^{-1},$$

gdzie $\mathbf{u}'C_j = \{\mathbf{u}'\mathbf{x}_{j1}, \dots, \mathbf{u}'\mathbf{x}_{jn_j}\}$, $MAD(Y) = \operatorname{med}\{|Y - \operatorname{med}(Y)|\}$.

Dla ustalonego zbioru danych (rozkładu prawdopodobieństwa) głębia projekcyjna w \mathbb{R}^p jest afinicznie niezmiennicza, quasi-wypukła, głębia punktu

zmierza do zera, jeśli norma punktu zmierza do nieskończoności, głębia przyjmuje maksimum w centrum symetrii rozkładu.

Własności klasyfikatora projekcyjnego porównano z liniową oraz kwadratową funkcją dyskryminacyjną oraz z klasyfikatorem głębii Tukey'a na przykładzie znanego zbioru danych (Fisher, 1936; Rao, 1982), dotyczącego 3 gatunków Irysa, rozpatrywanych ze względu na 4 cechy kwiatu. Rozpatrywano próby uczące wielkości 25:25:25 oraz 40:40:40. Na ich podstawie klasyfikowano obserwacje należące do zbioru Irys. Na dobre własności proponowanej reguły wskazują wyniki zawarte w tabeli 1.

Tabela 1

Wyniki badań symulacyjnych proponowanego klasyfikatora projekcyjnego

	3 × 25 obserwacji w próbie uczącej			
	liniowa funkcja dyskryminacyjna	kwadratowa funkcja dyskryminacyjna	klasyfikator głębii projekcyjnej	klasyfikator głębii Tukey'a
Rzeczywisty błąd predykcji	2.6%	4%	3.3%	66%
	3 × 40 obserwacji w próbie uczącej			
	Rzeczywisty błąd predykcji	3.8%	2.6%	2.6%

Generowano także 00 razy zbiór 3000 dwuwymiarowych obserwacji : 1000 z rozkładu Marshalla-Olkina, 1000 z izotropowego rozkładu normalnego, 1000 ze skośnego rozkładu Studenta *T*. Ze zbioru pobierano 100 × 100 × 100 próby uczące. Na jej podstawie klasyfikowano zbiory 3000 obserwacji. Na dobre własności proponowanej reguły wskazują wyniki zawarte w tabeli 2.

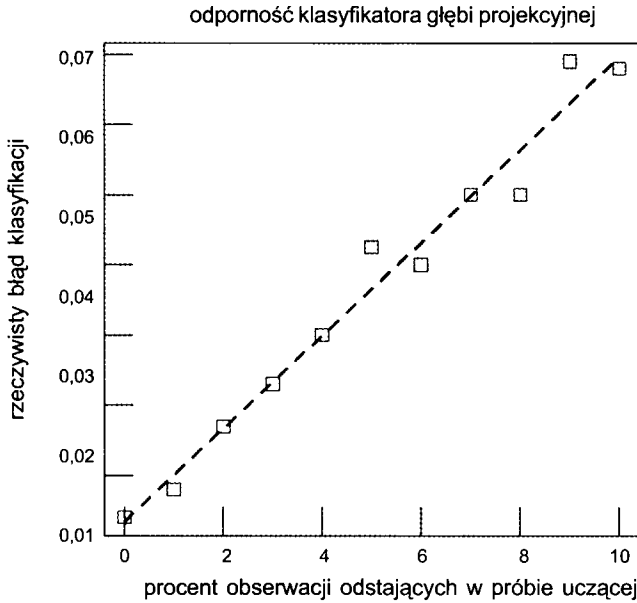
Tabela 2

Wyniki badań symulacyjnych proponowanego klasyfikatora projekcyjnego

	Klasyfikator			
	liniowa funkcja dyskryminacyjna	kwadratowa funkcja dyskryminacyjna	klasyfikator głębii projekcyjnej	klasyfikator głębii Tukey'a
Rzeczywisty błąd predykcji	12.6%	12.6%	0.3%	68%

W celu oszacowania punktu załamania próby skończonej generowano zbiory 3000 dwuwymiarowych obserwacji z populacji będącej mieszaniną trzech sko-

śnych rozkładów T -Studenta o równych udziałach a różniących się parametrami położenia i kształtu. Następnie zastępowano po 0%, 1%, ..., 10% obserwacji w 3×100 elementowej próbie uczącej, reprezentującej każde skupisko obserwacjami dalece odstającymi od centrów rozkładów mieszaniny. Obliczano rzeczywisty błąd klasyfikacji po takim zastąpieniu.



Ryc. 9. Punkt załamania proponowanej reguły dyskryminacyjnej

Propozycja reguły grupowania obserwacji na jednorodne skupiska

Przypuśćmy, że dysponujemy n obserwacjami p wymiarowymi $C_0 = \{x_1, \dots, x_n\}$. Naszym celem jest wskazanie pewnego optymalnego rozbitcia zbioru C_0 na k jednorodnych rozłącznych podzbiorów p wymiarowych C_1, \dots, C_k , $k \geq 2$, $C_i \cap C_j = \emptyset$, $i \neq j$, $\bigcup C_i = C_0$.

Propozycja reguły. Niech $\tilde{C}_1, \dots, \tilde{C}_k$, $k \geq 2$, $\tilde{C}_i \cap \tilde{C}_j = \emptyset$, $i \neq j$, $\bigcup C_i = C_0$, będzie pewnym możliwym rozbitciem zbioru obserwacji C_0 . Powiemy, że rozbitcie $\tilde{C}_1, \dots, \tilde{C}_k$ jest lepsze niż rozbitcie trywialne C_0 i \emptyset , jeżeli zachodzi:

$$vol(D_{PRO}^{\alpha}(C_0)) > \sum_{i=0}^k vol(D_{PRO}^{\alpha}(\tilde{C}_i)), \text{ dla ustalonego } \alpha \in 0, 1, \quad (18)$$

gdzie $vol(D_{PRO}^{\alpha}(C_0))$ oznacza objętość centralnego obszaru centralnego.

5. ODPORNE WYKRYWANIE PUNKTU ZMIANY TENDENCJI W WYBRANYCH MODELACH REGRESJI

Zagadnienie wskazania punktu zmiany tendencji (ang. *change point*) w modelu regresji było studiowane w ekonomii m.in. w kontekście badań zmian natury zależności pomiędzy zjawiskami powyżej pewnego poziomu jednego z nich. Zagadnienie wiąże się m.in. z badaniem zysku z inwestycji z wielkością poniesionych wydatków, badaniem przyrostu naturalnego i PKB na mieszkańca. Na poziomie technicznym pojawia się w zagadnieniach dyskryminacji i klasyfikacji obiektów, ma związek z jakością tychże metod.

W literaturze statystycznej przedstawiono szereg podejść dotyczących estymacji i weryfikacji hipotez dla modeli regresji w dwóch fazach.

Chen (1998) zaproponował wykorzystanie kryterium informacyjnego Schwarza (SIC), do wskazania punktu zmiany tendencji w modelu liniowym przy założeniu normalności.

Osorio i Galea (2005) rozszerzyli wyniki Chena (1998) na model regresji liniowej z niezależnymi błędami *t*-Studenta.

Sformułowanie problemu

Opierając się na n niezależnych obserwacji $(Y_1, x_1^t), (Y_2, x_2^t), \dots, (Y_n, x_n^t)$, zamierzamy zweryfikować hipotezę, że parametry regresji nie zmieniają się:

$$H_0 : Y_i = x_i^t \beta + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

przeciw hipotezie alternatywnej, że następuje zmiana (parametrów regresji) na nieznaną pozycję k , nazywaną punktem zmiany tendencji:

$$H_1 : Y_i = x_i^t \beta_1 + \varepsilon_i, \quad i = 1, 2, \dots, k,$$

$$Y_i = x_i^t \beta_2 + \varepsilon_i, \quad i = k + 1, \dots, n,$$

gdzie $\beta_1 = (\beta_0, \beta_1, \dots, \beta_{p-1})^t$, $\beta_2 = (\beta_0^*, \beta_1^*, \dots, \beta_{p-1}^*)^t$, i ε oznacza błąd.

W celu rozwiązania powyższego problemu Osorio i Galea (2005), odwołując się do pracy Chena (1998), proponują zamienić proces testowania hipotez na procedurę wyboru modelu z wykorzystaniem kryterium informacyjnego Schwarza (SIC) definiowanego:

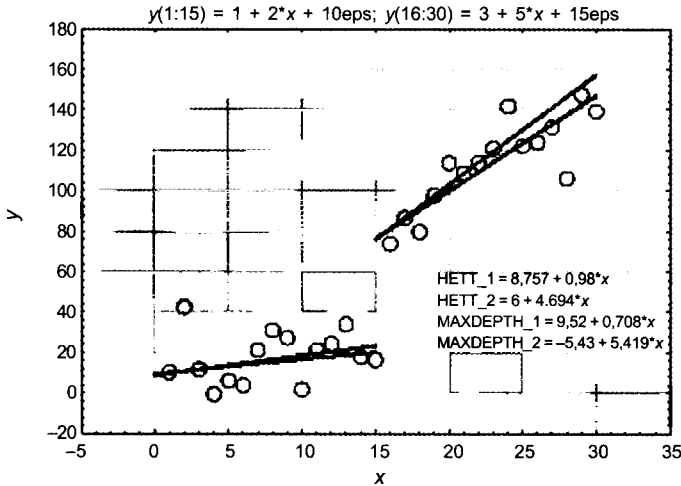
$$SCI = -2L(\hat{\theta}) + s \log n, \quad (19)$$

gdzie $L(\theta)$ to logarytm wiarygodności obliczony dla oszacowań parametrów uzyskanych metodą największej wiarygodności, s jest liczbą parametrów w modelu, n to wielkość próby.

Osorio i Galea rozważają model regresji liniowej:

$$Y_i = x_i^t \beta + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (20)$$

gdzie $x_i, i = 1, 2, \dots, n$ odpowiada i -temu wierszowi $n \times p$ macierzy X , $\beta = (\beta_0, \beta_1, \dots, \beta_{p-1})^t$ jest wektorem nieznanych parametrów, natomiast ε_i są to niezależne błędy losowe o takim samym rozkładzie $t(0, \phi, n)$.



Ryc. 10. Ilustracja procedury detekcji punktu zmiany tendencji

W celu estymacji parametrów i obliczania informacji Schwarza wykorzystują zmodyfikowany algorytm EM.

Przy założeniu H_0 , nie występuje zmiana współczynników regresji; przy założeniu H_1 mamy zbiór możliwych modeli ze zmianami w punkcie p or $p + 1$ or...or $n - p$.

Obliczamy wartość SIC przy założeniu H_0 oraz przy założeniu H_1 .

Wybieramy model z punktem zmiany na pozycji k , jeżeli dla pewnego k :

$$SIC(n) > SIC(k). \quad (21)$$

W przypadku gdy hipoteza zerowa jest odrzucana, estymator największej wiarygodności punktu zmiany tendencji oznaczany przez k , powinien spełniać warunek:

$$SIC(k) = \min\{SIC(k) : p \leq k \leq n - p\} = \max\{L_k(\theta) : p \leq k \leq n - p\}. \quad (22)$$

w dalszej części porównujemy propozycję z podejściem kryterium informacyjnego Schwarza wykorzystując pakiet autorstwa Juliana Taylora heteroscedastic t -regression package (R Project).

Propozycja

Uporządkujmy obserwacje według wartości zmiennej x :

$$(y, x_{(1)}), \dots, (y, x_{(p-1)}), (y, x_{(p)}), (y, x_{(p+1)}), \dots, (y, x_{(n-1)}), (y, x_{(n)}),$$

dla $p = \left\lceil \frac{4n}{10} \right\rceil, \dots, \left\lceil \frac{6n}{10} \right\rceil$ („okno poszukiwań długości $2/10 \cdot n$ ”).

Oznaczmy:

$$\mathbf{Z}^n = (y, x_{(1)}), (y, x_{(2)}), \dots, (y, x_{(n-1)}), (y, x_{(n)}),$$

$$\mathbf{Z}_-^n = (y, x_{(1)}), (y, x_{(2)}), \dots, (y, x_{(p-1)}), (y, x_{(p)}),$$

$$\mathbf{Z}_+^n = (y, x_{(p)}), (y, x_{(p+1)}), \dots, (y, x_{(n-1)}), (y, x_{(n)}),$$

Następnie obliczmy:

1. Estymator maksymalnej głębi regresyjnej dla wszystkich obserwacji $T_r^*(\mathbf{Z}^n) = (b_0, b_1) = \hat{b}_r$ oraz głębię maksymalnego dopasowania $rdepth(\hat{b}_r)$.

2. Estymator maksymalnej głębi regresyjnej dla $T_r^*(\mathbf{Z}_-^n) = (b_0^-, b_1^-) = \hat{b}_r^-$ oraz głębię tego dopasowania $rdepth(\hat{b}_r^-)$.

3. Estymator maksymalnej głębi regresyjnej dla $T_r^*(\mathbf{Z}_+^n) = (b_0^+, b_1^+) = \hat{b}_r^+$ oraz głębię tego dopasowania $rdepth(\hat{b}_r^+)$.

Jeżeli dla pewnego $p = \left\lceil \frac{4n}{10} \right\rceil, \dots, \left\lceil \frac{6n}{10} \right\rceil$ ma miejsce:

$$rdepth(\hat{b}_r) < \min\{rdepth(\hat{b}_r^-), rdepth(\hat{b}_r^+)\}, \quad (23)$$

wtedy uznajemy, że w punkcie p następuje zmiana parametrów regresji. W takim przypadku za parametry regresji dla obserwacji $i = 1, 2, \dots, p$ przyjmujemy \hat{b}_r^- natomiast dla obserwacji $i = p + 1, \dots, n$ — za parametry przyjmujemy \hat{b}_r^+ .

Z przeprowadzonych przez autora badań symulacyjnych wynika, że w przypadku nie występowania punktu zmiany tendencji propozycja spisuje się lepiej niż metoda kryterium Schwarza w przypadkach, gdy błędy mają rozkład normalny bądź Cauchy'ego.

W przypadku, gdy błąd ma rozkład normalny oraz występują dwie obserwacje odstające kryterium SIC sprawuje się nieco lepiej.

W przypadku występowania punktu zmiany tendencji obie rozważane metody zachowują się podobnie, metoda SIC odznacza się nieco mniejszym rozrzutem wskazania.

6. ANALIZA DANYCH PANELOWYCH Z WYKORZYSTANIEM GŁĘBI REGRESYJNEJ

W klasycznej analizie regresji na ogół zakłada się, że obserwacje są pobierane z tej samej populacji, są niezależne i o takim samym rozkładzie. W przypadku analizy regresji prowadzonej z wykorzystaniem modeli mieszanych stosuje się słabsze założenia. Mianowicie, dane mogą tworzyć skupiska, obserwacje pomiędzy skupiskami są niezależne, jednak nie muszą być niezależne wewnątrz skupisk.

Modele mieszane wydają się użyteczne np. w badaniach gmin z uwzględnieniem podziału na województwa, w badaniu wydatków konsumpcyjnych z uwzględnieniem grupy dochodowej itd.

W ekonometrii przyjęto nazywanie obserwacji danymi panelowymi wówczas, gdy dotyczą poszczególnych jednostek przekrojowych w dłuższym czasie (więcej niż jednym okresie).

Przypuśćmy, że na dane patrzymy z punktu widzenia liniowego modelu mieszanego, w postaci zaproponowanej przez Lairda i Ware'a w 1982 roku:

$$y_i = x_i\beta + Z_i b_i + \varepsilon_i, \quad i = 1, \dots, N, \quad (24)$$

gdzie:

y_i — jest $n_i \times 1$ wektorem odpowiedzi i -tego skupiska (n_i odpowiedzi jednostek z i -tego skupiska);

x_i — jest $n_i \times m$ macierzą ustalonych efektów w i -tym skupisku;

β — uśredniony dla wszystkich skupisk wektor parametrów związanych ze stałymi efektami;

Z_i — $n_i \times k$ macierz eksperymentu efektów losowych w i -tym skupisku;

ε_i — $n_i \times 1$ wektor błędu dla i -tego skupiska, wektor o niezależnych składowych, każda o przeciętnej zero i wariancji σ^2 ;

b_i — jest $k \times 1$ wektorem parametrów związanych z efektami losowymi w i -tym skupisku, wektor o zerowej przeciętnej i macierzy kowariancji $D^* = \sigma^2 D$.

Zakładamy, że macierz $\sum X_i^t X_i$ jest nieosobliwa oraz, że $\sum n_i > m$ dla zapewnienia identyfikowalności modelu (24) względem β . Dla zapewnienia identyfikowalności modelu (24) względem σ^2 i D , zakładamy, że przynajmniej jedna macierz $z_i^t z_i$ jest dodatnio określona oraz, że $\sum_{i=1}^N (n_i - k) > 0$.

Wypada wspomnieć także, że często wykorzystuje się skalowaną macierz kowariancji efektów losowych:

$$D = \frac{1}{\sigma^2} D^* = \frac{1}{\sigma^2} Cov(b_i). \quad (25)$$

W celu estymacji parametrów modelu (24) metodą NW zakładamy, że:

$$\varepsilon_i \sim N(0, \sigma^2 I_{n_i}), \quad b_i \sim N(0, \sigma^2 D). \quad (26)$$

Warto zauważyć, że przy założeniach (3) model (1) można zapisać w następującej postaci brzegowej:

$$y_i \sim N(X_i \beta, \sigma^2 (I_{n_i} + Z_i D Z_i^t)), \quad i = 1, \dots, N. \quad (27)$$

Okazuje się, że ustalając macierz D , logarytm wiarygodności dla modelu (24) maksymalizowany jest przez uogólniony estymator najmniejszych kwadratów NK

$$\hat{\beta}_{UNK} = \left[\sum_{i=1}^N X_i^t (I + Z_i D Z_i^t) \right]^{-1} \left[\sum_{i=1}^N X_i^t (I + Z_i D Z_i^t)^{-1} y_i \right], \quad (28)$$

Zauważmy, że w specjalnym przypadku, gdy $D = 0$, estymator (36) sprowadza się do zwykłego estymatora NK:

$$\hat{\beta}_{NK} = (\sum X_i^t X_i)^{-1} (\sum X_i^t y_i). \quad (26)$$

W wielu zastosowaniach wykorzystuje się szczególną wersję liniowego modelu mieszanego (24), w którym dopuszcza się jeden efekt losowy dotyczący wyrazu wolnego, a który definiowany jest jako:

$$y_{ij} = a_i + \gamma^t x_{ij} + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, N, \quad (27)$$

gdzie y_{ij} interpretowane jest jako j -ta odpowiedź w i -tym skupisku na wartość predyktora x_{ij} , $\gamma = [\alpha \beta]$, przy czym to uśredniony dla całej populacji wyraz wolny, β uśredniony dla całej populacji parametr nachylenia.

W modelu (27) wyraz wolny w skupisku i interpretowany jest jako suma uśrednionego dla całej populacji parametru α oraz efektu losowego specyficznego dla i -tego skupiska:

$$a_i = \alpha + b_i. \quad (28)$$

Na ogół zakłada się, że $\varepsilon_{ij} \sim N(0, \sigma^2)$ oraz $b_i \sim N(0, \sigma^2 d)$ są niezależne, gdzie σ^2 jest wariancją błędu wewnątrz skupiska oraz d jest skalowaną wariancją efektu losowego.

Model (27) może pojawić się przykładowo w następującej sytuacji. Rozważamy grupę N województw w ujęciu powiatów. Każde województwo badamy ze względu na stopę bezrobocia, przeciętne wynagrodzenie, saldo migracji w gminach województwa, przyrost naturalny. Przypuśćmy, że interesuje nas zależność przyrostu naturalnego od pozostałych zmiennych. Kluczowym dla stosowalności modelu (27) w przedstawionej sytuacji jest założenie, że struktura zależności nie zmienia się od województwa do województwa, co znaczy, że γ jest ustalonym wektorem.

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

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

Eksperymenty symulacyjne wskazują na niską efektywność estymatora UNK w przypadkach, gdy wariancja efektów losowych jest istotnie większa od wariancji błędu oraz gdy wariancja efektów losowych lub wariancji błędu są nieznane.

Propozycja odpornego estymatora głębi regresyjnej

Przypuśćmy, że na obserwacje patrzymy z punktu widzenia modelu (27).

Oznaczmy przez $Z_i^{n_i}$ zbiór par obserwacji y i x w i -tym skupisku tzn.

$$Z_i^{n_i} = ((x_{i1}, y_{i1})^t (x_{i2}, y_{i2})^t \dots (x_{in_i}, y_{in_i})^t)^t, \quad i = 1, \dots, N.$$

Niech $T_i^{rdepth}(Z_i^{n_i}) = \operatorname{argmax}_{rdepth}(\mathbf{b}, Z_i^{n_i}) = \hat{\mathbf{b}}_i = (\hat{b}_i^0, \hat{b}_i^1)$ oznacza estymator maksymalnej głębi regresyjnej wektora parametrów $\mathbf{b}_i = (b_i^0, b_i^1)$ liniowej funkcji regresji $y_i = (b_i^0, b_i^1)x$ w i -tym skupisku.

Niech:

$$T^{rdepth}(((Z_i^{n_i})^t \dots (Z_N^{n_N})^t)^t) = \operatorname{argmax}_{rdepth}(\mathbf{b}, ((Z_i^{n_i})^t \dots (Z_N^{n_N})^t)^t) = \hat{\mathbf{b}} = (\hat{b}_0, \hat{b}_1), \quad (29)$$

oznacza estymator maksymalnej głębi regresyjnej wektora parametrów $\hat{\mathbf{b}} = (\hat{b}_0, \hat{b}_1)$ liniowej funkcji regresji $y = b_0 + b_1x$ dla obserwacji z wszystkich skupisk.

Weźmy za estymatory parametrów modelu zdefiniowanego przez (27) i (28)

$\hat{\beta}^{rdepth} = \hat{b}_1$ — nachylenie uśrednione dla wszystkich skupisk;

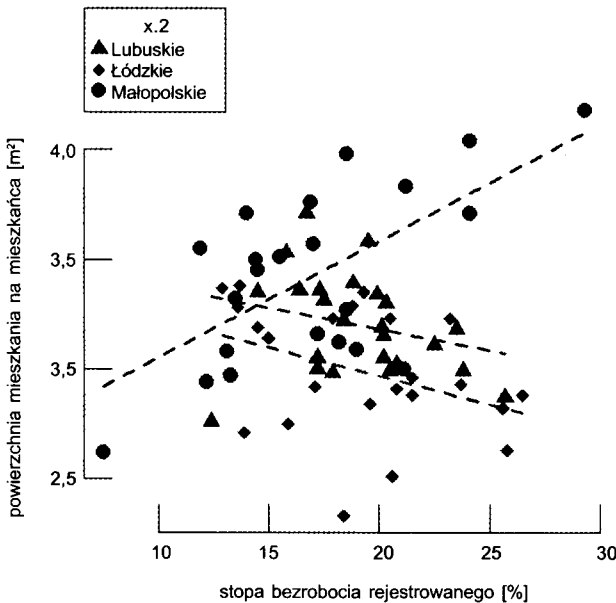
$\hat{\beta}^{rdepth} = \hat{b}_0$ — wyraz wolny uśredniony dla całej populacji;

$\hat{\alpha}_i^{rdepth} = \hat{b}_{i0}$ — wyraz wolny specyficzny dla i -tego skupiska, $i = 1, \dots, N$;

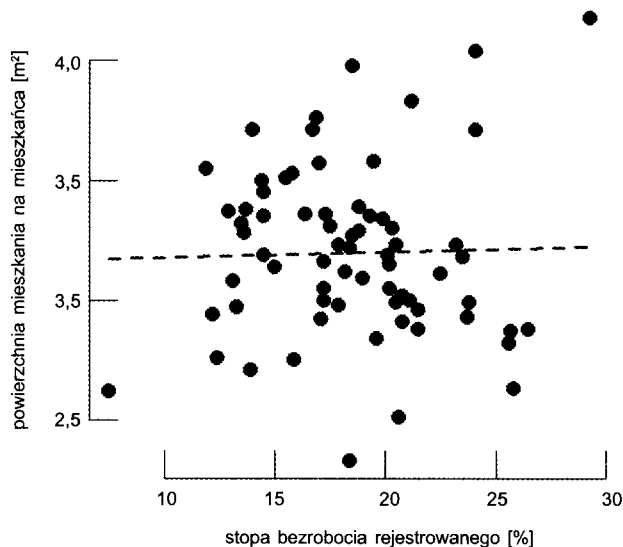
$b_i^{rdepth} = \alpha_i^{rdepth} - \alpha^{rdepth}$ efekt losowy w i -tym skupisku, $i = 1, \dots, N$.

Rozpatrzmy 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 m²) w roku 2005.

Zauważmy, że zaletą danych panelowych jest możliwość weryfikacji oraz złagodzenia założeń, które domyślnie są przyjmowane w analizie danych przekrojowych. Przy stosunkowo łagodnych założeniach dotyczących media-

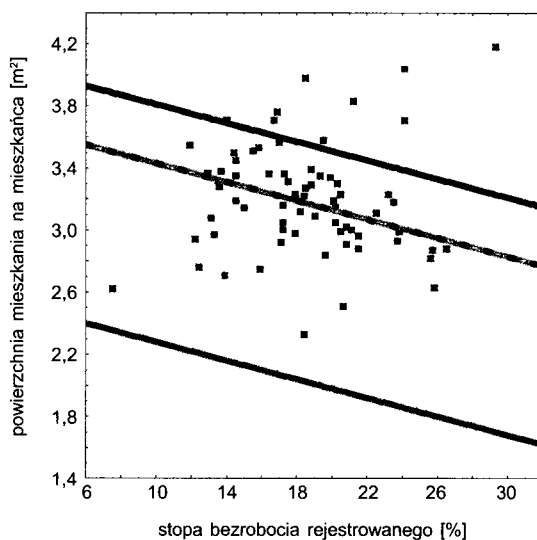


Ryc. 11. Powierzchnia mieszkania vs. stopa bezrobocia — dane dotyczące powiatów trzech województw traktowanych oddzielnie

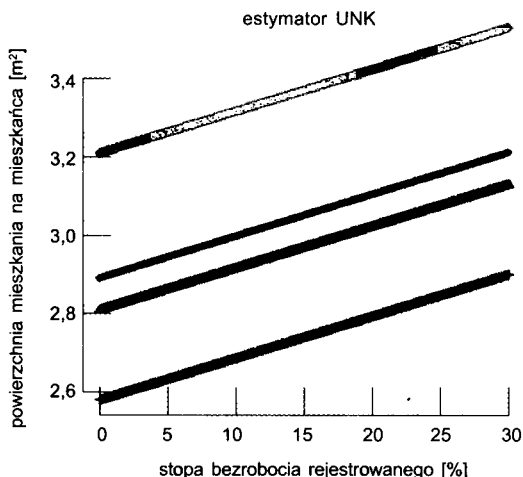


Ryc. 12. Powierzchnia mieszkania vs. stopa bezrobocia — dane dotyczące powiatów trzech województw traktowanych łącznie

ny 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.



Ryc. 13. Powierzchnia mieszkania vs. stopa bezrobocia — dane dotyczące powiatów w trzech województwach — estymator maksymalnej głębi parametrów modelu



Ryc. 14. Powierzchnia mieszkania vs. stopa bezrobocia — dane dotyczące powiatów w trzech województwach — estymator UNK parametrów modelu

Estymator maksymalnej głębokości regresyjnej dobrze radzi sobie ze skośnymi oraz heteroskedastycznymi rozkładami błędów i efektów losowych. Wyniki symulacji wskazują na nieobciążoność i niezłą efektywność estymatora maksymalnej głębokości regresyjnej w porównaniu z uogólnionym estymatorem NK oraz niezależnymi ocenami parametrów w skupiskach za pomocą zwykłego estymatora NK.

7. PODSUMOWANIE

Poszukiwanie nieparametrycznych i odpornych zarazem procedur statystycznych, adekwatnych dla badania wielowymiarowych układów społeczno-ekonomicznych jest ważne zarówno z teoretycznych, jak i praktycznych względów. Zdaniem autora naszkicowana w pracy perspektywa badań, ma zastosowanie do lepszego zrozumienia struktury współzależności układów ekonomicznych. Przedstawione własności proponowanych metod, odwołujących się do koncepcji głębokości danych, wydają się wystarczającym uzasadnieniem dla dalszych studiów nad tym zagadnieniem.

BIBLIOGRAFIA

- Azzalini A., Capitanio A. 2003. *Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution*, J. Roy. Statist. Soc. B 65, 367–389.
- Chaudhuri P. 1996. *On a Geometric Notion of Quantiles for Multivariate Data*, Journal of the American Statistical Association 91, 862–872.

- Chen J. 1998. *Testing for a Change Point in Linear Regression Models*, Communications in Statistics — Theory & Methods 27, 2481–2493.
- Demidenko E. 2004. *Mixed Models — Theory and Applications*, John Wiley & Sons, Inc., Hoboken–New Jersey.
- Dyckerhoff R. 2004. *Data Depths Satisfying the Projection Property*, Allgemeines Statistisches Archiv 88, 163–190.
- Hoberg R., Mosler K. 2003. *Classification based on data depth*, Bulletin of the ISI 54th Session.
- Hoberg R., Mosler K. 2006. *Data analysis and classification with the zonoid depth*, [w:] *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, R. Liu, R. Serfling, D. Souvaine red., American Mathematical Society, 2006, 49–59.
- Koltchinskii V. 1997. *M-estimation, Convexity and Quantiles*, The Annals of Statistics 25, 435–477.
- Kosiorowski D. (w druku). *Odporność metod klasyfikacyjnych wykorzystujących funkcje głębi*, Acta Universitatis Lodzianensis, Folia Oeconomica, Materiały z Konferencji Multivariate Statistical Analysis 2007.
- Kosiorowski D. (w druku). *Analiza danych panelowych z wykorzystaniem głębi regresyjnej*, Acta Universitas Lodzianensis, Folia Oeconomica, Materiały z Konferencji SPSG'07.
- Kosiorowski D. 2007. *O Kwantylovym funkcjonele asymetrii rozkładu wektora losowego w badaniach szeregów finansowych*, [w:] *Dynamiczne modele ekonometryczne*, Z. Zieliński red., Wydawnictwo UMK w Toruniu, Toruń, 129–136.
- Kosiorowski D. 2007. *O odpornej analizie regresji w ekonomii na przykładzie koncepcji głębi regresyjnej*, Przegląd Statystyczny 1, 109–121.
- Kosiorowski D. 2007. *Krzywa skali — odporna i nieparametryczna metoda badania rozrzutu wektora losowego i stopnia zależności jego rozkładów brzegowych*, Ryzyko Asekuracja Statystyka nr 44, Raport Techniczny Katedry Statystyki AE we Wrocławiu, 35–36.
- Kosiorowski D. 2008. *About Robust Detection of a Change — Point in Selected Linear Regression Models*, Konferencja Multivariate Statistical Analysis 2006, Uniwersytet Łódzki, Acta Universitatis Lodzianensis, Folia Oeconomica 216, 109–117.
- Kosiorowski D. 2008. *Krzywa skali — odporna i nieparametryczna metoda badania rozrzutu wektora losowego i stopnia zależności jego rozkładów brzegowych*, Konferencja Statystyka Aktuarialna, Teoria i Praktyka, Wrocław, Badania Operacyjne i Decyzyjne 4, 47–60.
- Krzyśko M. (w druku). *Modele Klasyfikacyjne*, referat plenarny na konferencji Multivariate Statistical Analysis 2006, Łódź.
- Laird N., Ware J. 1982. *Random-effects models for longitudinal data*, Biometrics 38, 963–974.
- Liu R.Y., Parelius J.M., Singh K. 1999. *Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference (with discussion)*, The Annals of Statistics 27, 783–858.
- Maddala G.S. 2006. *Ekonometria*, PWN, Warszawa.
- Mizera I. 2002. *On Depth and Deep Points: A Calculus*, Annals of Statistics, 30, 1681–1736.
- Mizera I., Muller Ch.H. 2002. *Breakdown Points of Cauchy Regression-Scale Estimators*, Statistics and Probability Letters 57, 79–89.
- Mosler K. 2002. *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*, Springer, New York.
- Osorio F., Galea M. 2005. *Detection of a Change — Point in Student-t Linear Regression Models*, Statistical Papers 45, 31–48.
- Romanazzi M. 2004. *Data Depth and Correlation*, Allgemeines Statistisches Archiv 88, 191–214.
- Rousseeuw P.J., Hubert M. 1998. *Regression Depth*, J. Amer. Statist. Assoc. 94, 388–433.
- Rousseeuw P.J., Leroy A.M. 1987. *Robust Regression and Outlier Detection*, Wiley, New York.
- Serfling R.J. 2004. *Nonparametric Multivariate Descriptive Measures Based on Spatial Quantiles*, Journal of Statistical Planning and Inference 123, 259–278.
- Serfling R.J. 2004. *Nonparametric Multivariate Descriptive Measures Based on Spatial Quantiles*, Journal of Statistical Planning and Inference 123, 259–278.

- Serfling R.J. 2006. *Multivariate Symmetry and Asymmetry*, [w:] *Encyclopedia of Statistical Sciences*, wyd. 2, S. Kotz, N. Balakrishnan, C.B. Read, B. Vidakovic red., Wiley, 5338–5345.
- Van Aelst S., Rousseeuw P.J. 2000. *Robustness Properties of Deepest Regression*, J. Multiv. Analysis 73, 82–106.
- Wang J., Serfling R. 2006. *Influence Functions for a General Class of Depth — Based Generalized Quantile Functions*, Journal of Multivariate Analysis 97, 810–826.
- Zuo Y. 2003. *Projection Based Depth Functions and Associated Medians*, The Annals of Statistics 31(5), 1460–1490.
- Zuo Y., Serfling R. 2000. *General Notions of Statistical Depth Function*, The Annals of Statistics 28, 461–482.
- Zuo Y. 2004. *Robustness of Weighted L_p — Depth and L_p — Median*, AStA 88, 215–234.
- Zuo Y. 2003. *Projection Based Depth Functions and Associated Medians*, The Annals of Statistics 31(5), 1460–1490.
- Zuo Y., Cui H., Young D. 2004. *Influence Function and Maximum Bias of Projection Depth Based Estimators*, The Annals of Statistics 32(1), 189–218.
- Zuo Y., Cui H., Young D. 2004. *Influence Function and Maximum Bias of Projection Depth Based Estimators*, The Annals of Statistics 32(1), 189–218.
- Zuo Y., Serfling R. 2000. *Nonparametric Notions of Multivariate „Scatter Measure“ and „More Scattered“ Based on Statistical Depth Function*, Journal of Multivariate Analysis 75, 62–78.

STOPY ZWROTU A WIELKOŚĆ OBROTÓW NA GPW W WARSZAWIE

HENRYK GURGUL

Katedra Ekonomii i Ekonometrii
Akademia Górniczo-Hutnicza
PL 30-059 Kraków, ul. Gramatyka 10
e-mail: h.gurgul@neostrada.pl

TOMASZ WÓJTOWICZ

Katedra Ekonomii i Ekonometrii
Akademia Górniczo-Hutnicza
PL 30-059 Kraków, ul. Gramatyka 10
e-mail: twojtow@agh.edu.pl

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie 14 października 2008 przez autorów.

ABSTRACT

H. Gurgul, T. Wójtowicz. *Returns versus trading volume on Warsaw Stock Exchange*, Folia Oeconomica Cracoviensia 2008–2009, 49–50: 31–45.

In the paper the results of empirical investigations of dynamic relationships between extreme trading volume and subsequent stock returns on Warsaw Stock Exchange are presented. The event study methodology is applied. The dynamic relationship between the financial variables is rather weak and depends on kind and size of the stock exchange. The high-volume-return-premium is more pronounced for small size stocks with lower liquidity levels.

KEY WORDS — SŁOWA KLUCZOWE

stock exchange, returns, trading volume, dynamic relationships
giełdy, stopy zwrotu, wielkość obrotów, zależności dynamiczne

1. WSTĘP

W literaturze finansowej znane są przykłady badań dotyczących relacji pomiędzy stopami zwrotu a odpowiadającą im wielkością obrotów, zarówno gdy chodzi o pojedyncze spółki, jak i cały rynek — należą do nich m.in. prace Lorente i in. (2002); Gerlacha i in. (2006); Lee i Rui (2006); McMillana (2007) i inne.

Jeśli założy się, że rynki kapitałowe są efektywne informacyjnie, to zjawiska takie jak krótkoterminowa tendencja do kontynuacji, jeśli chodzi o stopy, zwrotu czy długoterminowa tendencja tych stóp do powrotu do stanu równowagi (ang. *mean reverting process*) bądź znaczenie wielkości obrotów w prognozowaniu stóp zwrotu i na odwrót są ciekawymi zjawiskami, które wymagają wyjaśnienia. Niektórzy autorzy, tacy jak: deBondt i Thaler (1985, 1987, 1990); Cambell i in. (1993); Conrad i in. (1994) oraz Hong i Stein (1999) przypuszczają, że stwierdzona przez nich empirycznie ujemna autokorelacja stóp zwrotu wynika z początkowej ich nadreakcji na nową, ważną dla rynku kapitałowego informację.

Z dostępnych w literaturze wyników badań wzajemnej zależności pomiędzy wielkością obrotów a stopami zwrotu wynika, że to raczej stopy zwrotu wpływają na wolumen (np. Chen i in., 2001; Lee i Rui, 2002; Hiemstra i Jones, 1994). Zależność przeciwna, zwłaszcza liniowa, jest mniej powszechna. Liniowy wpływ wolumenu na stopy zwrotu obserwowany jest głównie na rynkach wschodzących, natomiast na rynkach dojrzałych wpływ ten ma charakter nieliniowy — wolumen oddziałuje głównie na zmienność stóp zwrotu. Wspomniane powyżej wyniki dotyczą sytuacji ogólnej, od której w szczególnych wypadkach notowane są odstępstwa. W związku z tym szczególnego znaczenia nabiera badanie występujących na rynku anomalii, które mogą zostać wykorzystane przez inwestorów.

Aby stwierdzić, że ma miejsce zjawisko dążenia cen lub stóp zwrotu akcji do długoterminowych cen lub stóp równowagi, trzeba uwzględnić dłuższy horyzont czasowy. W krótkim czasie obserwuje się często, że po dodatnich (ujemnych) stopach zwrotu następują dodatnie (ujemne) stopy zwrotu w dniach następnych. Piszą o tych zjawiskach m.in. Chan i Hamed (2000). Są i tacy autorzy, jak Chincarini i Lorente-Alvarez (1999), Lee i Swaminathan (2000) czy Connolly i Stivers (2003), którzy nawet chwilowe efekty wiązali z wpływem innych czynników, takich jak wielkość obrotów, i właśnie ta wielkość była kluczową w ich strategiach inwestycyjnych na rynkach akcji.

Bardzo ważnym kierunkiem badawczym w ostatnim okresie jest kierunek związany z teorią tzw. nadmiernej pewności inwestora (ang. *overconfidence*). Należy on do dziedziny finansów behawioralnych, a zatem wiąże się z psychologią rynków finansowych (Dreman i Lufkin, 2000; Huddart i in., 2005). Głównym założeniem tej teorii jest stwierdzenie, że niektórzy inwestorzy przeceniają dokładność swoich informacji. Zgodnie z tą teorią wszystkich inwesto-

rów można podzielić na racjonalnych i tych, którzy przeceniają precyzję swoich informacji. Uwzględniając przesłanki psychologiczne Daniel i in. (1997) oraz Odean (1998) przyjmują, że nieracjonalni inwestorzy, czyli inwestorzy należący do drugiej z wymienionych grup, charakteryzują się specyficznym zachowaniem na giełdzie, a przede wszystkim: po pierwsze — sprzedają lub kupują więcej niż racjonalni inwestorzy, po drugie — handlują na giełdzie bardziej agresywnie, co zwiększa zmienność kursów akcji, po trzecie — inwestorzy ci wykazują nadmierną reakcję nawet w odniesieniu do swoich prywatnych informacji, co prowadzi do niewłaściwej wyceny walorów, po czwarte — anomalie, które wynikają z takiego zachowania inwestorów, są niwelowane w dłuższej perspektywie czasowej.

Oprócz badań wpływu wielkości obrotów na wielkość stóp zwrotu (Aggarwal i Sun, 2003) badano też efekt odwrotny, w ramach znanego w literaturze finansowej tzw. efektu dyspozycji. W wypadku tego efektu zakłada się, że inwestorzy mają tendencję do stosunkowo szybkiej realizacji zysków na poziomie ostatnio zrealizowanej dodatniej stopy zwrotu. Inwestorzy ci zwlekają jednak ze sprzedażą w sytuacji ujemnych stóp zwrotu (oczekują na poprawę). Zarówno teoria nadmiernej pewności inwestora, jak i efekt dyspozycji prowadzą do wniosku o dodatnim związku pomiędzy dodatnimi stopami zwrotu a wielkością obrotów.

W pewnych sytuacjach te dwa efekty mogą się nakładać, tzn. wysokie obroty mogą wynikać z nadmiernego zaufania inwestorów do posiadanych informacji, jak i z faktu występowania w poprzednich dniach dodatnich stóp zwrotu. Wydaje się, że kwantyfikacja obu efektów nie jest łatwym zagadnieniem. W ostatnich latach pojawiły się prace, np. Statman i in. (2004), Glaser i Weber, (2005), Chuang i Lee (2006), których celem było sprawdzenie teorii o nadmiernym zaufaniu inwestora do posiadanych informacji. Autorzy zbadali m.in. z punktu widzenia indywidualnych inwestorów zachowanie kilku tysięcy ich portfeli.

Istotną pracą nawiązującą do roli wielkości obrotów na rynkach akcji jest praca Gervaisa i in. (2001). Z badań tych autorów wynika, że po bardzo dużych (małych) wielkościach obrotów obserwuje się wysokie (niskie) stopy zwrotu, jeśli za punkt odniesienia przyjąć wielkość stóp zwrotu w warunkach przeciętnej aktywności handlowej inwestorów na giełdzie. Autorzy określają tę zależność mianem premii wysokiego wolumenu (ang. *high-volume return premium*). W praktyce podkreśla się, że zarówno w okresie wysokich, jak i niskich stóp zwrotu obserwuje się wzrost aktywności handlowej. Ten ostatni wypadek może wynikać stąd, że inwestorzy starają się usunąć ze swoich portfeli silnie zniżkujące akcje.

Gervais i in. (2001) przypisują istnienie premii wysokiego wolumenu zmianom w postrzeganiu akcji danej spółki przez inwestorów. Nagły wzrost wolumenu, a w zasadzie informacja o tym fakcie, powoduje, że inwestorzy do

tej pory niezainteresowani spółką zaczynają ją dostrzegać. Powoduje to wzrost liczby potencjalnych nabywców jej akcji. Jednak z powodu ograniczonej liczby akcji spółki, będących na rynku konsekwencją tego zainteresowania, jest wzrost ceny. Autorzy wykazują ponadto, że wpływ ekstremalnych wartości wolumenu na stopy zwrotu nie zależy od innych czynników, takich jak: ceny akcji, zapowiedzi wypłat dywidendy, ogłoszenia wyników spółki czy też płynności akcji.

Dalsze badanie wpływu ekstremalnych wielkości obrotów na stopy zwrotu było prowadzone m.in. przez Kaniela i in. (2003, 2005), Aggarwala i Suna (2003), Lei i Li (2006).

Celem naszego artykułu jest przedstawienie wyników badania wpływu ekstremalnych wartości wolumenu na stopy zwrotu spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie. Oprócz zagadnienia istnienia wpływu — bądź nie — podane zostaną wyniki badań nad możliwością praktycznego zastosowania ewentualnych, wykrytych zależności. W dalszej części pracy przedstawimy najpierw źródła danych i ich charakterystykę. Następnie omówimy zastosowaną w pracy metodologię, a szczególnie wykorzystaną wersję analizy zdarzeń, będącą podstawowym wykorzystanym narzędziem badawczym. W dalszej części zaprezentujemy i omówimy wyniki obliczeń na podstawie danych dla GPW w Warszawie. Pracę zakończy krótkie podsumowanie najważniejszych wyników badań.

2. DANE

Obliczenia zostały przeprowadzone na podstawie danych obejmujących ciągle dzienne stopy zwrotu oraz wolumen 73 spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie w systemie notowań ciągłych co najmniej w okresie od 2 stycznia 2001 do 30 września 2007, zarówno na rynku podstawowym, jak i równoległym oraz wolnym. W tej grupie znajdują się również dane dotyczące akcji Narodowych Funduszy Inwestycyjnych. Dane pochodzą z Ceduły GPW oraz ogólnodostępnych portali internetowych bossa.pl i rynek.bph.pl.

W celu zmniejszenia skośności i kurtozy danych dotyczących wolumenu zastosowano do nich transformację logarytmiczną.

3. METODOLOGIA

Badania przeprowadzone w tej pracy oparte są na analizie zdarzeń. Ponieważ głównym celem jest zbadanie wpływu ekstremalnych (tzn. bardzo dużych lub bardzo małych) wartości wolumenu na stopy zwrotu, interesującym zdarzeniem będzie wystąpienie bardzo wysokich lub też bardzo niskich wartości

wolumenu. Nasza definicja ekstremalnych wartości wolumenu jest zbliżona do występującej w pracy Gervaisa i in. (2001). Przyjmujemy, że spółka w danym dniu ma bardzo wysoki (niski) wolumen, jeżeli jest on większy (mniejszy) niż wolumen odnotowany w ciągu 50 wcześniejszych notowań oraz w ciągu tych 50 wcześniejszych sesji nie występuje tak zdefiniowany bardzo wysoki ani bardzo niski wolumen. To dodatkowe założenie jest konieczne, by możliwe było zastosowanie analizy zdarzeń. Chodzi zwłaszcza o jednoznaczną identyfikację dnia zajścia tak zdefiniowanego zdarzenia. Przy badaniu wpływu wolumenu na stopy zwrotu należy bowiem zadbać, by zdarzenia były od siebie odizolowane i nie zakłócały wzajemnie swoich oddziaływań. Oczywiście, nie trudno wyobrazić sobie sytuację, gdy w jakimś dniu spółka odnotowała zdefiniowany powyżej bardzo wysoki wolumen, a w dniu następnym jeszcze wyższy. Jednak jako dzień zajścia zdarzenia zostanie zakwalifikowany do próby tylko pierwszy z wymienionych dni. Dodanie założenia o odizolowaniu zdarzeń powoduje duże zmniejszenie ich liczby. Dla rozważanych danych jest to redukcja z 2822 do 258 w przypadku bardzo wysokiego wolumenu i z 3039 do 317 w przypadku bardzo niskiego wolumenu.

Przyjęta definicja zdarzenia pozwala na oddzielne badanie wpływu niskiego i wysokiego wolumenu. Jak wynika z pracy Gervaisa i in. (2001) wpływ ten jest odmienny. Bardzo wysoki wolumen powoduje występowanie dodatnich stóp zwrotu akcji, natomiast bardzo niski wolumen powoduje występowanie ujemnych stóp zwrotu. Siła wpływu ekstremalnych wartości wolumenu zależy również od wielkości spółki — słabnie ona wraz ze wzrostem kapitalizacji spółki. W związku z tym, analogicznie jak Gervais i in. (2001), spółki w poszczególnych latach zostały podzielone na cztery grupy w zależności od kapitalizacji na koniec roku poprzedniego. Grupę najmniejszych spółek stanowiło 20% spółek o najniższej kapitalizacji, następne 30% stanowi grupę spółek małych, kolejne 30% to grupa spółek średnich, a 20% spółek o najwyższej kapitalizacji na koniec poprzedniego roku stanowi grupę spółek dużych. W kolejnym roku kalendaryzowym klasyfikacja spółek do poszczególnych grup jest dokonywana ponownie.

Powyższe definicje pozwoliły na zaliczenie każdego zdarzenia do jednej z ośmiu klas: wysoki lub niski wolumen spółek: bardzo małych, małych, średnich i dużych.

4. ANALIZA ZDARZEŃ

Jak zostało wcześniej wspomniane, do oceny wpływu ekstremalnych wartości wolumenu na stopy zwrotu spółki zastosowano metodologię analizy zdarzeń. Jeżeli dzień zdarzenia oznaczymy jako $t = 0$, to jako okno estymacyjne przyjęto dane z notowań dla dni $t = -52, \dots, -3$, natomiast jako okno zdarzenia przy-

jęto dane dla dni $t = -2, \dots, 5$. Na podstawie stóp zwrotu spółki R_t , oraz indeksu WIG oznaczonych przez M_t , a pochodzących z okna estymacyjnego (przed zdarzeniem), wyestymowano parametry modelu rynkowego:

$$AR_t = R_t - \alpha - \beta M_t, \quad (1)$$

dla $t = -52, \dots, 5$.

Model ten posłużył do obliczenia zwykłych stóp zwrotu AR_t w oknie estymacyjnym i w oknie zdarzenia. Następnie dla każdej klasy, zarówno dla okna estymacyjnego, jak i okna zdarzenia, zdefiniowane zostały przekrojowe, średnie, zwykłe stopy zwrotu \overline{AR}_t :

$$\overline{AR}_t = \frac{1}{N} \sum_{i=1}^N AR_{i,t}, \quad (2)$$

gdzie N stanowi licznosc danej klasy zdarzeń.

Odchylenie standardowe z próby średnich zwykłych stóp zwrotu oblicza się według wzoru:

$$\hat{\sigma}[\overline{AR}_t] = \sqrt{\frac{1}{49} \sum_{t=-52}^{-3} (\overline{AR}_t - \overline{\overline{AR}}_t)^2}, \quad (3)$$

przy czym $\overline{\overline{AR}}_t = \frac{1}{50} \sum_{t=-52}^{-3} \overline{AR}_t$ jest przeciętną zwykłą stopą w oknie estymacyjnym.

W celu weryfikacji hipotezy o braku wpływu zdarzenia na stopy zwrotu w dniu t stosuje się statystyki:

$$t_R = \frac{\overline{AR}_t}{\hat{\sigma}[\overline{AR}_t]}. \quad (4)$$

Przy założeniu, że średnie zwykłe stopy zwrotu mają rozkład normalny zdefiniowane statystyki mają rozkład t -Studenta o $N-1$ stopniach swobody.

5. WYNIKI EMPIRYCZNE

Dla rozważanych spółek w okresie od 2 stycznia 2001 do 30 września 2007 wyodrębniono 258 zdarzeń polegających na wystąpieniu bardzo wysokiego wolumenu oraz 317 wypadków — bardzo niskiego wolumenu. W każdej w tych grup przeprowadzono badanie wpływu odnotowanych ekstremalnych wielkości obrotów na stopy zwrotu. Wyniki testowania istotności średnich zwykłych stóp zwrotu w oknie zdarzenia zostały zebrane w tabeli 1.

Jak można zauważyć, w wypadku występowania bardzo wysokiego wolumenu, średnie zwykłe stopy zwrotu w dniach najbliższych zdarzeniu są dodatnie oraz istotne. Dotyczy to ponadprzeciętnych, średnich stóp zwrotu

Tabela 1

Dzienne średnie zwykłowe stopy zwrotu

Dzień t	Bardzo wysoki wolumen			Bardzo niski wolumen		
	\overline{AR}_t (w %)	stat. t	p -value	\overline{AR}_t (w %)	stat. t	p -value
-2	0,024	0,173	0,863	0,027	0,229	0,819
-1	0,409**	2,896	0,004	-0,211	-1,805	0,072
0	2,212**	15,663	0,000	-0,232*	-1,993	0,047
+1	0,763**	5,402	0,000	-0,161	-1,381	0,168
+2	0,337*	2,389	0,018	-0,047	-0,405	0,686
+3	0,212	1,504	0,134	-0,221	-1,895	0,059
+4	0,251	1,777	0,077	-0,283*	-2,425	0,016
+5	-0,234	-1,654	0,099	-0,024	-0,205	0,838

Źródło: obliczenia własne.

Objaśnienia: * — wartość istotna na poziomie 5%, ** — wartość istotna na poziomie 1%.

jeden dzień przed zdarzeniem, w dniu zdarzenia i w ciągu dwu kolejnych dni po zajściu zdarzenia. Oznacza to, że bardzo wysokim wartościom wolumenu towarzyszą zwiększone stopy zwrotu akcji. Jednak z drugiej strony, istotna wartość średniej zwykłowej stopy zwrotu w dniu poprzedzającym zdarzenie świadczy o tym, że to raczej wzrost cen akcji powoduje wzrost zainteresowania inwestorów transakcjami kupna/sprzedaży, co prowadzi do wzrostu wolumenu. Jest to potwierdzenie wspomnianego wcześniej efektu dyspozycji w zakresie dotyczącym zachowań inwestorów, gdy stopy zwrotu na giełdzie są wysokie. Z kolei wzrost zainteresowania inwestorów, a więc też popytu na akcje, może powodować wzrost ich kursów, co oznacza utrzymywanie się w następnych dniach dodatnich zwykłych stóp zwrotu. Natomiast bardzo niskiemu wolumenowi towarzyszą ujemne wartości zwykłych stóp zwrotu, przy czym istotne są one tylko w dniu zdarzenia. Występowanie ujemnych stóp zwrotu przed i w samym dniu ekstremalnie niskiej wielkości obrotów może potwierdzać efekt dyspozycji w części dotyczącej wstrzymywania się inwestorów ze sprzedawaniem akcji, gdy sprzedaż wiąże się ze stratami (ujemne stopy zwrotu).

Podsumowując można stwierdzić, że wyniki przeprowadzonego badania, a w szczególności istotne dodatnie zwykłowe stopy zwrotu w następnym dniu po zdarzeniu, potwierdzają występowanie na GPW premii wysokiego wolumenu. Jednak pełniejszy obraz wpływu ekstremalnych wartości wolumenu na stopy zwrotu na GPW daje analiza uwzględniająca wielkość spółek. Jej wyniki zostały zebrane w tabeli 2.

Dzienne, średnie, zwykłe stopy zwrotu spółek o różnej wielkości
(bardzo małe, małe, średnie i duże)

Bardzo wysoki wolumen								
Dzień t	bardzo małe spółki (41 zdarzeń)		małe spółki (69 zdarzeń)		średnie spółki (87 zdarzeń)		duże spółki (61 zdarzeń)	
	\overline{AR}_t (w %)	stat. t	\overline{AR}_t (w %)	stat. t	\overline{AR}_t (w %)	stat. t	\overline{AR}_t (w %)	stat. t
-2	0,007	-0,015	-0,338	-1,017	0,145	0,657	0,275	1,625
-1	1,641**	3,028	0,256	0,873	0,356	1,616	-0,170	-0,038
0	4,477**	7,510	3,293**	11,071	1,421**	6,458	0,596**	3,341
+1	1,157*	2,479	1,386**	4,721	0,499*	2,267	0,170	0,654
+2	1,330*	2,541	0,414	1,507	0,245	1,115	-0,285	-1,880
+3	0,725	1,380	0,372	1,194	0,025	0,114	-0,045	-0,284
+4	-0,099	-0,149	0,585*	2,011	0,496*	2,252	-0,241	-1,379
+5	-1,440**	-2,775	0,206	0,658	-0,047	-0,212	-0,187	-1,212
Bardzo niski wolumen								
Dzień t	bardzo małe spółki (65 zdarzeń)		małe spółki (97 zdarzeń)		średnie spółki (88 zdarzeń)		duże spółki (67 zdarzeń)	
	\overline{AR}_t (w %)	stat. t	\overline{AR}_t (w %)	stat. t	\overline{AR}_t (w %)	stat. t	\overline{AR}_t (w %)	stat. t
-2	0,275	0,679	-0,149	-0,538	0,036	0,240	0,027	0,375
-1	0,135	0,333	-0,702*	-2,545	-0,172	-0,771	0,115	0,534
0	-0,449	-1,108	-0,169	-0,614	-0,186	-0,756	-0,175	-0,833
+1	-0,234	-0,577	-0,337	-1,221	0,007	-0,008	-0,058	-0,279
+2	-0,103	-0,255	-0,013	-0,045	0,118	0,487	-0,260	-1,110
+3	-0,145	-0,357	-0,615*	-2,228	-0,067	-0,315	0,072	0,447
+4	-0,576	-1,421	-0,050	-0,180	-0,379	-1,622	-0,210	-1,041
+5	-0,210	-0,518	0,069	0,251	0,102	0,362	-0,143	-0,550

Źródło: obliczenia własne.

Objaśnienia: * — wartość istotna na poziomie 5%, ** — wartość istotna na poziomie 1%.

Wynika z nich, że wypadkom występowania bardzo wysokiego wolumenu towarzyszą istotnie dodatnie zwyżkowe stopy zwrotu niezależnie od wielkości spółek. Tylko w przypadku najmniejszych spółek można odnotować istotne statystycznie, dodatnie zwyżkowe stopy zwrotu występujące dzień wcześniej. Fakt ten jest potwierdzeniem przypuszczenia, że w przypadku małych spółek o niskiej płynności wzrost ceny akcji może bardziej niż w przypadku dużych spółek skłaniać ich właścicieli do realizacji zysków. Z drugiej strony, na GPW w Warszawie istotną rolę odgrywają inwestorzy indywidualni, w tym drobni, którzy podejmują decyzje inwestycyjne na podstawie obserwacji ruchów cen poszczególnych akcji (ang. *noisy traders*). Małe spółki nie są obiektem szczególnego zainteresowania analityków giełdowych czy mediów. Inwestorzy instytucjonalni, posiadający zwykle własnych profesjonalnych analityków i dostęp do mediów, inwestują zwykle w większe spółki, których akcje są wybierane do portfela na podstawie szczegółowej, profesjonalnej analizy obecnej sytuacji spółki i jej perspektyw rozwojowych. Popyt na akcje małych spółek, generowany przez drobnych inwestorów podejmujących decyzje inwestycyjne na podstawie obserwacji wzrostu cen, może prowadzić, i — jak potwierdzają nasze wyniki empiryczne — prowadzi do wzrostu obrotów w następnym dniu (dniu zdarzenia). Wzrost obrotów prowadzi do istotnego wzrostu kursu akcji, a więc i stóp zwrotu w dniach następnych. Można więc mówić o samonapędzającym się mechanizmie: ceny akcji rosną, więc inwestorzy (zwłaszcza drobni) je kupują, wzmożony popyt na akcje powoduje dalszy wzrost ich cen itd. Ten proces nie trwa jednak w nieskończoność. Na rynku z dużym udziałem inwestorów indywidualnych przysłowiowa „zapałka” (niekiedy nawet tylko płotka) może spowodować „jazdę w dół”. Z naszych badań wynika także wniosek, że obecność dodatnich zwyżkowych stóp zwrotu dzień przed wystąpieniem bardzo wysokiego wolumenu (odnotowana w tab. 1) może być spowodowana głównie zachowaniem inwestorów posiadających akcje najmniejszych spółek.

Biorąc pod uwagę cel pracy, najważniejsze jest stwierdzenie występowania istotnie większych od zera zwyżkowych stóp zwrotu w następnym dniu po wystąpieniu bardzo wysokiego wolumenu. Tak jest w przypadku wszystkich spółek z wyjątkiem największych. Należy podkreślić, że w przypadku małych spółek zwyżkowe stopy zwrotu są istotne na poziomie 1%. Oznacza to, że w przypadku większości spółek (szczególnie małych) wystąpienie bardzo dużego wolumenu powoduje w szczególności wzrost stóp zwrotu w dniu następnym. Trochę inna sytuacja ma miejsce w przypadku występowania bardzo niskich wartości wolumenu. Z przeprowadzonego badania wynika, że tego typu zdarzeniom nie towarzyszy istotna zmiana stóp zwrotu. Należy jednak podkreślić, że średnie zwyżkowe stopy zwrotu są w większości ujemne.

Z praktycznego punktu widzenia ciekawe jest czy potwierdzony powyżej wpływ wysokiego wolumenu na stopy zwrotu akcji na GPW może zostać

wykorzystany do konstrukcji zyskowej strategii inwestycyjnej. W tym celu zbadane zostały stopy zwrotu spółek w trakcie kolejnych notowań po zdarzeniu. Rozważone zostały dyskretne stopy zwrotu obliczone w stosunku do ceny zamknięcia poprzedniego notowania, czyli $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$ lub w stosunku do ceny zamknięcia notowania w dniu zdarzenia $\tilde{R}_t = \frac{P_t - P_0}{P_0}$. Ten drugi przypadek ma pomóc w wyborze optymalnego czasu sprzedaży waloru zakupionego w dniu zdarzenia.

Tabela 3

Średnie stopy zwrotu (w %) w kolejnych dniach po zdarzeniu (zdarzeniem był ekstremalnie wysoki wolumen) liczone względem wcześniejszego notowania oraz względem ceny zamknięcia w dniu zdarzenia

	Dzień t	Spółki				
		wszystkie	bardzo małe	małe	średnie	duże
R_t	$t = 1$	0,96*	1,72	1,41*	0,66	0,38
	$t = 2$	0,51	1,53	0,68	0,41	-0,20
	$t = 3$	0,24	0,55	0,31	0,04	0,26
	$t = 4$	0,45*	0,43	0,63*	0,56*	0,09
	$t = 5$	-0,20	-1,19*	0,27	-0,03	-0,35
\tilde{R}_t	$t = 1$	0,96*	1,72	1,41*	0,66	0,38
	$t = 2$	1,52*	3,56	2,10*	1,10	0,17
	$t = 3$	1,87*	4,74	2,46*	1,12	0,43
	$t = 4$	2,18*	4,33	3,08*	1,68*	0,52
	$t = 5$	1,95*	3,01	3,37*	1,60*	0,18

Źródło: obliczenia własne.

Objaśnienia: * — wartość istotna na poziomie 5%.

W większości przypadków uzyskane średnie stopy zwrotu są nieujemne, ale też są nieistotne. Występująca nieistotność średnich stóp zwrotu wynika zapewne w głównej mierze z małej liczności danych w poszczególnych grupach spółek. Stąd najczęściej istotnie różnych od zera średnich stóp zwrotu występuje, gdy rozważane są wszystkie spółki. W szczególności należy podkreślić istotność średnich stóp zwrotu obliczanych względem dnia zdarzenia. Wynik ten sugeruje wybór strategii polegającej na kupnie waloru w dniu zdarzenia i jego sprzedaży cztery dni później. Jeżeli pominięte zostaną koszty transakcyjne to taka strategia przynosi przeciętnie dla wszystkich spółek prawie 2,2%

zysku. Podobna sytuacja ma miejsce, gdy rozważania zostaną zawężone do grupy małych spółek, przy czym w tym przypadku oczekiwany zysk po pięciu dniach jest istotnie większy — wynosi prawie 3,4%. Dla pozostałych grup spółek nie można statystycznie potwierdzić występowania przesłanek przemawiających na korzyść takiej strategii.

Można się też zastanawiać czy uzyskane wyniki mogą być poprawione jeżeli uwzględnimy ewentualne dodatkowe czynniki, jak np. wartość stóp zwrotu w dniu zdarzenia, w okresie przed zdarzeniem lub też wartość wolumenu. W związku z tym w dalszym badaniu jako czynniki, które mogą determinować zachowanie się stóp zwrotu akcji w kolejnych notowaniach po dniu zdarzenia przyjęto: średnią stóp zwrotu z okna przed zdarzeniem, medianę stóp zwrotu z okna przed zdarzeniem medianę zwykłych stóp zwrotu w oknie przed zdarzeniem, stopę zwrotu akcji w dniu zdarzenia oraz zwykłą stopę zwrotu w dniu zdarzenia. W każdym z przypadków osobno badano spółki, dla których rozważane kryterium jest większe bądź mniejsze od zera. Średnie stopy zwrotu dla takich podgrup zostały zebrane w tabelach 4

Tabela 4

Średnie stopy zwrotu (w %) liczone względem dnia zdarzenia (zdarzeniem był ekstremalnie wysoki wolumen) w trakcie kolejnych notowań po zdarzeniu obliczone dla wszystkich spółek spełniających wybrane kryteria

Dzień t	Okno przed zdarzeniem			Dzień zdarzenia	
	\bar{R}_t	me R_t	me AR_t	R_0	AR_0
większe od 0:					
$t = 1$	1,13*	1,05*	0,90*	1,47*	1,46*
$t = 2$	2,05*	1,43*	1,10*	2,39*	2,29*
$t = 3$	2,33	1,62*	1,37*	2,92*	2,73*
$t = 4$	2,31*	1,95*	2,23*	3,13*	2,87*
$t = 5$	2,21*	1,61	1,99*	2,89*	2,64*
mniejsze od 0:					
$t = 1$	0,76*	1,21*	1,00*	-0,15	-0,13
$t = 2$	0,92	2,68	1,81*	-0,41	-0,13
$t = 3$	1,34*	3,89	2,20*	-0,51	0,01
$t = 4$	2,04*	3,99*	2,15*	-0,02	0,71
$t = 5$	1,66*	3,04*	1,92*	-0,02	0,48

Źródło: obliczenia własne.

Objaśnienia: * — wartość istotna na poziomie 5%, pogrubioną czcionką zaznaczono wartości większe niż średnie stopy zwrotu w tabeli 3.

Średnie stopy zwrotu i mediany (w %) małych spółek liczone względem dnia zdarzenia (zdarzeniem był ekstremalnie wysoki wolumen) w trakcie kolejnych notowań po zdarzeniu obliczone dla spółek spełniających wybrane kryteria

Dzień t	Okno przed zdarzeniem		Dzień zdarzenia		
	\bar{R}_t	me R_t	me AR_t	R_0	AR_0
większe od 0:					
$t = 1$	0,91	1,39	2,43*	1,70*	1,58*
$t = 2$	1,75	2,50	2,88*	2,46*	2,25*
$t = 3$	1,53	2,08	3,76*	2,39*	2,11*
$t = 4$	2,22	2,15	4,61*	2,83*	2,49*
$t = 5$	2,99*	1,81	4,71*	3,04*	2,86*
mniejsze od 0:					
$t = 1$	1,91*	1,47*	0,68	0,49	0,92
$t = 2$	2,46*	2,03*	1,56*	0,97	1,67*
$t = 3$	3,40*	2,46	1,54	2,64*	3,49*
$t = 4$	3,95*	3,34*	2,01	3,95*	4,79*
$t = 5$	3,75*	2,83*	2,42*	4,70*	4,86*

Źródło: obliczenia własne.

Oznaczenia: * — wartość istotną na poziomie 5%, pogrubioną czcionką zaznaczono wartości większe niż średnie stopy zwrotu w tabeli 3.

i 5 odpowiednio dla wszystkich spółek oraz tylko dla małych spółek. Te grupy zostały wybrane ponieważ głównie dla nich można oczekiwać istotnych zysków. Z kolei, w przeciwieństwie do tabeli 3, nie zaprezentowano wyników dla stóp zwrotu liczonych względem dnia poprzedniego, ponieważ w większości wypadków były one nieistotne.

Wyniki przeprowadzonych obliczeń wskazują, że zazwyczaj kupowanie wyłącznie akcji spółek, które w dniu zdarzenia odnotowały dodatnie (zwyżkowe) stopy zwrotu prowadzi do uzyskania istotnie dodatnich średnich stóp zwrotu, które ponadto są wyższe niż średnie stopy zwrotu obliczone bez uwzględnienia tego faktu. Jest to spowodowane występowaniem dodatniej korelacji pomiędzy wartością stóp zwrotu w dniu zdarzenia i w dniach kolejnych. Z drugiej strony ujemna korelacja pomiędzy stopami zwrotu w dniach po zdarzeniu a medianą zwykłych stóp zwrotu w oknie przed zdarzeniem również implikuje występowanie zwiększonych średnich stóp zwrotu, przy

czym w dwóch przypadkach są one nieistotne. Dodatkowo stopy zwrotu po dniu zdarzenia występują także w przypadku spółek, których mediana stopy zwrotu w dniu zdarzenia była ujemna.

Inaczej wygląda to w przypadku rozważenia wyłącznie małych spółek. Tu największe znaczenie zdaje się mieć ujemna korelacja stóp zwrotu w dniach po zdarzeniu ze średnią stopą zwrotu z okresu przed zdarzeniem oraz dodatnia korelacja z medianą zwykłych stóp zwrotu w tymże okresie. Kupno akcji małych spółek zgodnie z którymkolwiek z tych kryteriów (każde większe od zera) prowadzi do zwiększenia oczekiwanego zysku, przy czym najlepsze wyniki są w przypadku zastosowania podziału według mediany zwykłych stóp zwrotu.

Jak widać w drugiej części tabeli 5 w przypadku kupna akcji małych spółek, dla których wymienione wyżej parametry pozycyjne w oknie przed zdarzeniem lub w dniu zdarzenia były ujemne najwyższe stopy zwrotu z akcji gwarantują spółki, dla których średnia stopa zwrotu w oknie przed zdarzeniem była ujemna (istotność ponadprzeciętnych stóp zwrotu we wszystkich pięciu dniach po zdarzeniu). Wysokie stopy zwrotu, ale dopiero w dniach 3, 4 i 5 po zdarzeniu zapewniają także akcje spółek, które w dniu zdarzenia wykazywały ujemne stopy zwrotu lub ujemne ponadprzeciętne stopy zwrotu.

6. UWAGI KOŃCOWE

W pracy zostało przeprowadzone badanie wpływu ekstremalnych wartości wolumenu na stopy zwrotu spółek notowanych na GPW w Warszawie. Badanie to zostało przeprowadzone z wykorzystaniem metodologii analizy zdarzeń. Uzyskane wyniki świadczą o istnieniu dodatniego wpływu bardzo wysokiego wolumenu (nazywanego w tej pracy zdarzeniem) na stopy zwrotu w ciągu następných notowań. W większości spółek, zwłaszcza małych, stwierdzono bowiem występowanie istotnie dodatnich zwykłych stóp zwrotu w dniu po zdarzeniu. Uzyskane wyniki potwierdzają, występującą w literaturze hipotezę o tzw. premii wysokiego wolumenu i mogą zostać wykorzystane do budowy odpowiedniej strategii inwestycyjnej, którą opisano w artykule. W przypadku ekstremalnie małych obrotów (ten fakt był także uważany przez nas za zdarzenie) zwykłe stopy zwrotu były co prawda w większości ujemne, ale statystycznie nieistotne. Wyniki naszych badań empirycznych dostarczają też argumentów na rzecz efektu dyspozycji dla małych spółek z GPW w Warszawie. Istotna dodatnia wartość średniej zwykłej stopy zwrotu w dniu poprzedzającym zdarzenie świadczy bowiem o tym, że wzrost cen akcji powoduje wzrost zainteresowania inwestorów transakcjami kupna/sprzedaży, co prowadzi do wzrostu wielkości obrotów. Z drugiej strony, występowanie ujemnych stóp zwrotu przed i w samym dniu ekstremalnie niskiej wiel-

kości obrotów może potwierdzać efekt dyspozycji w części dotyczącej niechęci inwestorów do sprzedawania akcji, gdy sprzedaż oznaczałaby ich pogodzenie się ze stratami (ujemne stopy zwrotu).

Badania — przynajmniej pośrednio — przemawiają także za teorią nadmiernego zaufania części inwestorów do precyzji swoich informacji.

BIBLIOGRAFIA

- Aggarwal R., Sun M. 2003. *Trading volume extremes and the subsequent price behavior*, Financial Management Association, <http://207.36.165.114/Denver/Papers>.
- Campbell J., Grossman S., Wang J. 1993. *Trading volume and serial correlation in stock returns*, Quarterly Journal of Economics, 108, 905–939.
- Chan K., Hameed A., Tong W. 2000. *Profitability of momentum strategies in the international equity markets*, Journal of Financial and Quantitative Analysis, 35, 153–172.
- Chen G., Firth M., Rui O.M. 2001. *The Dynamic Relation Between Stock Returns, Trading Volume, and Volatility*. The Financial Review, 38, 153–174.
- Chincarini L.B., Lorente-Alvarez J.-G. 1999. *Volume and Return Information on Individual Stocks*, SSRN eLibrary.
- Chuang W.-I., Lee B.-S. 2006. *An empirical evaluation of the overconfidence hypothesis*, Journal of Banking and Finance, 30, 2489–2515.
- Connolly R., Stivers C. 2003. *Momentum and reversals in equity-index returns during periods of abnormal turnover and return dispersion*, The Journal of Finance, 58, 1521–1555.
- Conrad J., Hameed A., Niden C. 1994. *Volume and autocovariances in short-horizon individual security returns*, The Journal of Finance, 49, 1305–1329.
- Daniel K., Hirshleifer D., Subrahmanyam A. 1997. *A theory of overconfidence, self-attribution, and security market under- and overreaction*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2017&rec=1&srcabs=168472.
- deBondt W., Thaler R. 1985. *Does the stock market overreact?*, The Journal of Finance, 40, 793–805.
- deBondt W., Thaler R. 1987. *Further evidence on investor overreaction and stock market seasonality*, The Journal of Finance, 42, 557–581.
- deBondt W., Thaler R. 1990. *Do security analysts overreact?*, The American Economic Review, 80, 52–57.
- Dreman D., Lufkin E. 2000. *Investor overreaction: Evidence that its basis is psychological*, The Journal of Psychology and Financial Markets, 1, 61–75.
- Gerlach R., Chen C., Lin D., Huang M.-H. 2006. *Asymmetric responses of international stock markets to trading volume*, Physica A, 360, 422–444.
- Gervais S., Kaniel R., Mingelgrin D. 2001. *The high-volume return premium*, The Journal of Finance, 56, 877–919.
- Glaser M., Weber M. 2005a. *Overconfidence and trading volume*, Technical report, Swedish Institute for Financial Research.
- Glaser M., Weber M. 2005b. *Which past returns affect trading volume*, Technical report, Swedish Institute for Financial Research.
- Gurgul H., Mestel R., Schleicher C. 2003. *Stock market reactions to dividend announcements: Empirical evidence from the Austrian stock market*, Swiss Society for Financial Market Research, 17, 332–350.
- Hiemstra C., Jones J.D. 1994. *Testing for linear and nonlinear Granger causality in the stock price — volume relation*, Journal of Finance, 49, 1639–1664.
- Hong H., Stein J. 1999. *A unified theory of underreaction, momentum trading, and overreaction in asset markets*, The Journal of Finance, 54, 2143–2184.

- Huddart S., Yetman M., Lang M. 2005. *Psychological factors, stock price paths, and trading volume*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1118080.
- Kaniel R., Li D., Starks L. 2003. *The high volume return Premium and the investor recognition hypothesis: International evidence*, Working Paper.
- Kaniel, R., Li, D., Starks, L. 2005. *Investor visibility events: Cross-country evidence*, Working Paper.
- Lee B.S., Rui O.M. 2002. *The dynamic relationship between stock returns and trading volume: Domestic and cross-country evidence*, *Journal of Banking and Finance*, 26, 51–78.
- Lee C., Swaminathan B. 2000. *Price momentum and trading volume*, *The Journal of Finance*, 55, 2017–2069.
- Lei A., Li H. 2006. *Trading volume shocks and subsequent stock returns*, http://www.fma.org/SLC/Papers/FMA_TradingVolumeShocks.pdf.
- Llorente G., Michaely R., Saar G., Wang, J. 2002. *Dynamic volume-return relation of individual stock*, *The Review of Financial Studies*, 15, 1005–1047.
- McMillan D. 2007. *Non-linear forecasting of stock returns: Does volume help?*, *International Journal of Forecasting*, 36, 115–126.
- Odean T. 1998. *Volume, volatility, price, and profit when all traders are above average*, *Journal of Finance*, 53, 1887–1934.
- Statman M., Thorley S., Vorkink K. 2004. *Investor overconfidence and trading volume*, http://www.afa-jof.org/pdfs/2004program/UPDF/P463_Behavioral_Finance.pdf.

BAYESOWSKIE GRANICZNE FUNKCJE KOSZTU DLA SEKTORA DYSTRYBUCJI ENERGII

JACEK OSIEWALSKI

Katedra Ekonometrii i Badań Operacyjnych
Uniwersytet Ekonomiczny w Krakowie
PL 31-510 Kraków, ul. Rakowicka 27
oraz Krakowska Akademia im. Andrzeja Frycza Modrzewskiego
e-mail: eosiewa@cyf-kr.edu.pl

RENATA WRÓBEL-ROTTER

Katedra Ekonometrii i Badań Operacyjnych
Uniwersytet Ekonomiczny w Krakowie
PL 31-510 Kraków, ul. Rakowicka 27
e-mail: eewrobel@cyf-kr.edu.pl

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie w dniu 9 grudnia 2008 r.

ABSTRACT

Renata Wróbel-Rotter. *Bayesian frontier cost functions for the electricity distribution sector*. Folia Oeconomica Cracovienisa 2008-2009, 49-50: 47-69.

The paper discusses the application of the Bayesian stochastic frontier cost functions as a tool for assessing and comparing the cost efficiency of firms. The analysis is based on the micro-economic production theory and the concept of inefficiency that is related to the existence of inner mechanisms responsible for observing costs higher than indicated by theory, given technical and economic environment of firms. The stochastic cost frontier allows to estimate individual cost efficiency and to precisely decompose the observed cost into theoretical categories, such as minimal excessive or systematic cost, which play crucial role in assessing the economic behavior of firms. From the empirical perspective the most important is the minimal cost, which denotes the cost that is indispensable to achieve the observed production level given technology, input prices, cost effects of random variability. The Bayesian estimation of stochastic cost frontiers for panel data allows to treat a number of issues unsolved on the classical ground, such as stability of empirical results, assessment of the uncertainty of cost categories and efficiency coefficients. As a fundamental tool for the analysis we present

the Bayesian random effects model with constant efficiency distribution. The methodology is illustrated with the cost function for the 14 firms operating in the electricity distribution sector observed over 6 years.

KEY WORDS — SŁOWA KLUCZOWE

stochastic frontier models, cost efficiency, Bayesian econometrics
stochastyczne modele graniczne, efektywność kosztowa, ekonometria bayesowska

1. WSTĘP

Celem pracy jest prezentacja założeń i wyników analizy kosztów operacyjnych w sektorze dystrybucji energii w Polsce. Stosując bayesowskie wersje tzw. stochastycznych granicznych modeli kosztu dokonujemy oceny efektywności kosztowej wszystkich czternastu spółek dystrybucyjnych na podstawie danych rocznych z lat 2001–2006. Opis działania podmiotów gospodarczych, wykorzystany w badaniach, bazuje na ekonometrycznych modelach procesu produkcyjnego, które powstają w efekcie połączenia mikroekonomicznej teorii kosztu z rozważaniami uwzględniającymi nieefektywność jednostek produkcyjnych (przedsiębiorstw) na skutek niewłaściwego wykorzystania przez nie czynników produkcji

Graniczne modele produkcji lub kosztu były już szeroko wykorzystywane do analizy efektywności działania podmiotów, takich jak: linie lotnicze (Schmidt i Sickles, 1984), szpitale (Koop, Osiewalski i Steel, 1994, 1997), banki i ich oddziały (Marzec i Osiewalski 2003, 2008 oraz bibliografia tam zawarta), biblioteki akademickie i publiczne (Osiewalski i Osiewalska, 2003, 2006), elektrownie i elektrociepłownie (Wróbel-Rotter i Osiewalski, 2002; Wróbel-Rotter, 2004). Wiele zastosowań stochastycznych granicznych modeli produkcji i kosztu (niestandardowych i trudnych do analizy metodami niebayesowskimi) wykorzystuje możliwości wnioskowania bayesowskiego, wspomaganego numerycznie metodami Monte Carlo. Bayesowskie stochastyczne modele graniczne wprowadzili do ekonometrii van den Broeck, Koop, Osiewalski i Steel (1994) oraz Koop, Osiewalski i Steel (1994, 1997), którzy opracowali też metody Monte Carlo umożliwiające uzyskiwanie wyników empirycznych.

Bayesowski graniczny model kosztu zmiennego dla sektora dystrybucji energii został opracowany na zlecenie Urzędu Regulacji Energetyki (Osiewalski i Wróbel-Rotter, 2008), a proponowane metody i uzyskane wyniki stanowią alternatywę wobec tradycyjnych, wielce uproszczonych metod oceny efektywności kosztowej stosowanych uprzednio przez Urząd. Model przyjęty w pracy opisuje zależność kosztu operacyjnego (ponoszonego przez spółkę dystrybucyjną) od dużej liczby czynników o charakterze techniczno-ekonomicznym, które powinny determinować jego poziom, oraz od sposobu zarządzania, reprezentowanego przez zmienną nieobserwowalną (ukrytą) określającą efektywność. Podejście bayesowskie pozwala precyzyjnie określić niepewność

wnioskowania o wpływie wyróżnionych czynników oraz o niesprawności zarządzania (nieefektywności kosztowej). Ocena niepewności wnioskowania jest szczególnie ważna w przypadku małej liczby obserwacji, co ma miejsce w sektorze dystrybucji energii, w którym od 2001 roku działa 14 przedsiębiorstw. Poza samą dziedziną zastosowania, niniejsza praca różni się istotnie od dotychczasowych prac z zakresu bayesowskich granicznych modeli kosztu dzięki zdefiniowaniu nowych kategorii składowych kosztu i dokonaniu ich estymacji.

Następna (druga) część pracy poświęcona jest ogólnej prezentacji stochastycznego modelu granicznego i wskaźnika efektywności kosztowej. W części trzeciej definiujemy kategorie kosztu, jakie można teoretycznie rozważać (i empirycznie szacować) w ramach modelu granicznego. Część czwarta zawiera podstawowe informacje o bayesowskim ujęciu modelowania kosztu i nieefektywności kosztowej na podstawie danych panelowych. W części piątej przedstawiamy wyniki empiryczne, zwracając szczególną uwagę na wrażliwość ocen efektywności na dobór zmiennych objaśniających oraz na niepewność wnioskowania o składowych kosztu obserwowanego. Część szósta zawiera podsumowanie.

2. OGÓLNA CHARAKTERYSTYKA STOCHASTYCZNEGO MODELU GRANICZNEGO

Model przyjęty do oceny efektywności opisuje zależność obserwowanego kosztu zmiennego przedsiębiorstwa od czynników techniczno-ekonomicznych, kształtujących jego poziom, oraz od niesprawności zarządzania; jest on zapisywany w następującej ogólnej formie (por. Aigner i in., 1977; Meeusen i van den Broeck, 1977; Lovell, 1993; Greene, 1993):

$$C_{it}^{\text{obs.}} = \exp[f(x_{it}, \beta) + v_{it} + u_i]$$

a po obustronnym zlogarytmowaniu przyjmuje postać:

$$y_{it} = f(x_{it}, \beta) + u_i + v_{it}$$

gdzie y_{it} jest zmienną zależną, oznaczającą logarytm kosztu obserwowanego ($C_{it}^{\text{obs.}}$) w i -tym obiekcie ($i = 1, \dots, N$) w okresie t ($t = 1, \dots, T$), x_{it} jest wektorem egzogenicznych zmiennych objaśniających, $f(x_{it}, \beta)$ jest ogólnym oznaczeniem postaci analitycznej funkcji kosztu, β to wektor nieznanych parametrów tej funkcji, v_{it} to zmienna losowa o rozkładzie symetrycznym wokół zera (składnik czysto losowy), ujmująca wpływ czynników przypadkowych oraz błędu pomiaru kosztu, u_i oznacza zmienną losową przyjmującą wyłącznie wartości nieujemne i reprezentującą nieefektywność, suma $u_i + v_{it}$ jest złożonym składnikiem losowym. Zakładamy niezależność stochastyczną wszystkich zmiennych v_{it} i u_i , przyjmując dla v_{it} ten sam rozkład normalny, a dla u_i rozkłady wykładnicze.

W badaniu empirycznym przyjmujemy liniową względem parametrów postać analityczną funkcji f , tj. $f(x_{it}, \beta) = x_{it}\beta$, która może reprezentować model Cobba i Douglasa bądź translogarytmiczny. W przypadku prostszej funkcji Cobba i Douglasa (stosowanej w tej pracy ze względu na małą liczbę obserwacji i dużą liczbę zmiennych) wektor-wiersz x_{it} powinien zawierać logarytmy głównych zmiennych objaśniających krótkookresowy koszt zmienny, tj. wielkości produkcji, cen zmiennych czynników produkcji oraz nakładów czynników stałych. Postać liniowa dla logarytmów wszystkich zmiennych występujących w modelu kosztu jest opisem technologii dualnym wobec funkcji produkcji Cobba i Douglasa, stanowiąc równocześnie aproksymację pierwszego rzędu dla dowolnej gładkiej funkcji kosztu.

Stochastyczna graniczna funkcja kosztu zmiennego jest podstawą do konstrukcji miernika krótkookresowej nieefektywności kosztowej EK_{it} obiektu i w okresie t , określonej jako iloraz minimalnego kosztu zmiennego $C_{it}^{\min} = \exp[f(x_{it}, \beta) + v_{it}]$ (wynikającego z funkcji kosztu i wahań czysto losowych) do kosztu $C_{it}^{\text{obs}} = \exp(y_{it}) = \exp[f(x_{it}, \beta) + u_i + v_{it}]$ rzeczywiście poniesionego przez dany podmiot:

$$EK_{it} = \frac{\exp[f(x_{it}, \beta) + v_{it}]}{\exp[f(x_{it}, \beta) + v_{it} + u_i]} = \frac{C_{it}^{\min}}{C_{it}^{\text{obs}}} = \exp(-u_i),$$

przy czym $EK_{it} = EK_i$ ze względu na stałość u_i w czasie. Założenie w modelu dla danych przekrojowo-czasowych, że u_i jest efektem indywidualnym umożliwia precyzyjną estymację wskaźników efektywności, ponieważ szacując u_i wykorzystujemy obserwacje z kilku lat dla każdego z obiektów (a nie tylko jedną wartość, jak w przypadku danych przekrojowych). Konstrukcja wskaźnika efektywności powoduje, że zawiera się on w przedziale $(0, 1]$ i pozwala na dogodną interpretację: EK_i określa, jaka część kosztu poniesionego przez daną jednostkę w danym okresie jest kosztem uzasadnionym z ekonomicznego punktu widzenia, $(1-EK_i)$ wskazuje, jaka część jest kosztem nadwyżkowym, który mógłby zostać zredukowany.

3. KATEGORIE MODELOWE KOSZTÓW

Stochastyczna graniczna funkcja kosztu jest modelem strukturalnym, przedstawiającym koszt obserwowany i -tego podmiotu w okresie t jako iloczyn trzech czynników:

$$C_{it}^{\text{obs}} = \exp[f(x_{it}, \beta)] \cdot \exp(v_{it}) \cdot \exp(u_i),$$

z których pierwszy reprezentuje teoretyczny koszt graniczny (mikroekonomiczny), drugi — współczynnik zmiany kosztu na skutek uwarunkowań czysto losowych, a trzeci — stopień zwiększenia kosztu na skutek nieefektywno-

ści. Model strukturalny umożliwia definiowanie alternatywnych kategorii teoretycznych kosztu, których estymacja bayesowska (i ocena niepewności związanej z wnioskowaniem) jest w pełni możliwa poprzez brzegowe rozkłady *a posteriori*, a w uproszczonej postaci — poprzez wartości oczekiwane i odchylenia standardowe *a posteriori* kategorii składowych. W ramach rozważanego modelu definiujemy następujące teoretyczne kategorie kosztów:

1. Koszt graniczny (teoretyczny-mikroekonomiczny) *i*-tego obiektu w okresie *t*:

$$C_{it}^{gr} = \exp[f(x_{it}, \beta)],$$

który oznacza wielkość teoretyczną obliczoną na podstawie mikroekonomicznej funkcji kosztu bez uwzględnienia składnika czysto losowego i nieefektywności. Wielkość ta jest znaną funkcją parametrów, określającą teoretyczny koszt niezbędny do uzyskania danej wielkości produkcji, przy ustalonej technologii, cenach zmiennych czynników produkcji i nakładach czynników stałych, z pominięciem wpływu zakłóceń losowych i nieefektywności działania. Z formalnego punktu widzenia, nieznaną wartość kosztu granicznego jest znaną funkcją wektora parametrów strukturalnych β , więc jej rozkład *a posteriori* — o gęstości $p(C_{it}^{gr} | dane)$ — i jego charakterystyki można uzyskać z rozkładu *a posteriori* dla β .

2. Koszt systematyczny *i*-tego obiektu w okresie *t*:

$$C_{it}^{syst} = \exp[f(x_{it}, \beta) + u_i] = \exp[f(x_{it}, \beta)] \cdot \exp(u_i),$$

oznacza teoretyczny koszt mikroekonomiczny $\exp[f(x_{it}, \beta)]$, zwiększony o skutki nieefektywności, wolny natomiast od efektu zakłóceń losowych. Koszt systematyczny to koszt teoretyczny, jaki firma ponosiłaby przy danej technologii, zaobserwowanych poziomach zmiennych objaśniających i przy danym poziomie indywidualnej nieefektywności, ale bez zakłóceń przypadkowych. Różny od jeden iloraz wielkości kosztu obserwowanego i systematycznego jest wynikiem działania czynników określanych jako losowe, zawierających w rzeczywistości również wszystkie inne, drugorzędne wielkości, które nie zostały ujęte wśród zmiennych egzogenicznych przyjętych w modelu, a mają wpływ na kształtowanie się ponoszonego kosztu. Z tego względu dla niektórych przedsiębiorstw koszt obserwowany może być mniejszy od kosztu systematycznego, co oznacza, że czynniki inne niż ujęte w modelu umożliwiają rejestrowanie kosztu niższego niż wynika to z typowych uwarunkowań techniczno-ekonomicznych i danego, indywidualnego poziomu efektywności. Na mocy założeń strukturalnych, oddziaływanie czynników czysto losowych na koszt rzeczywisty jest niezależne od indywidualnej efektywności. Koszt systematyczny jest znaną funkcją zmiennej ukrytej u_i i wektora parametrów β , więc jego rozkład *a posteriori* o gęstości $p(C_{it}^{syst} | dane)$ można uzyskać z łącznego rozkładu *a posteriori* dla β i u_i .

3. Koszt niezbędny (minimalny) i -tego obiektu w okresie t :

$$C_{it}^{\min} = \exp[f(x_{it}, \beta) + v_{it}] = \exp[f(x_{it}, \beta)] \cdot \exp(v_{it}),$$

oznacza koszt teoretyczny obliczony z uwzględnieniem czynników losowych i pominięciem nieefektywności. Określa on minimalny koszt niezbędny do uzyskania obserwowanej wielkości produkcji przy danej technologii, ustalonych cenach czynników produkcji i nieprzewidywalnych dla firmy uwarunkowaniach zewnętrznych (czynnikach losowych). Stanowi podstawę do wyznaczenia wskaźnika indywidualnej krótkookresowej efektywności kosztowej EK_i , przedstawionego w poprzedniej części pracy. Gęstość rozkładu *a posteriori* kosztu niezbędnego $p(C_{it}^{\min} | dane)$ jest uzyskiwana z brzegowej gęstości *a posteriori* wskaźnika $EK_i = \exp(-u_i)$ przez jego przeskalowanie wielkością kosztu obserwowanego ($C_{it}^{\min} = C_{it}^{\text{obs}} \cdot EK_i$).

4. Koszt nadwyżkowy i -tego obiektu w okresie t :

$$C_{it}^{\text{nadw}} = C_{it}^{\text{obs}} - C_{it}^{\min} = C_{it}^{\text{obs}}(1 - EK_i),$$

określony jako różnica między kosztem obserwowanym a kosztem niezbędnym, wynikającym z mikroekonomicznej funkcji kosztu oraz zakłóceń losowych, jest wartością nieuzasadnioną, wynikającą wyłącznie z nieefektywnego działania.

Stosowane przez nas metody wnioskowania bayesowskiego pozwalają na prezentację niepewności związanej z estymacją danej kategorii kosztu za pomocą wykresu jej rozkładu *a posteriori* lub sumarycznie, poprzez tylko dwie liczby: wartość oczekiwaną *a posteriori* (ocenę punktową danej kategorii kosztu) i odchylenie standardowe *a posteriori* (bayesowski miernik błędu szacunku tej kategorii).

Proponowane kategorie modelowe kosztu wynikają bezpośrednio z alternatywnych dekompozycji kosztu obserwowanego na składowe będące wynikiem działania czynników systematycznych, losowych i nieefektywności; zachodzą następujące równości:

$$C_{it}^{\text{obs}} = C_{it}^{\text{gr}} + (C_{it}^{\text{obs}} - C_{it}^{\text{syst}}) + (C_{it}^{\text{syst}} - C_{it}^{\text{gr}}) = C_{it}^{\text{gr}} + RK1_{it} + RK4_{it},$$

$$C_{it}^{\text{obs}} = C_{it}^{\text{gr}} + (C_{it}^{\min} - C_{it}^{\text{gr}}) + (C_{it}^{\text{obs}} - C_{it}^{\min}) = C_{it}^{\text{gr}} + RK2_{it} + RK3_{it},$$

gdzie:

$C_{it}^{\text{syst}} = C_{it}^{\text{obs}} \exp(-v_{it})$, $C_{it}^{\min} = C_{it}^{\text{obs}} EK_i = C_{it}^{\text{gr}} \exp(v_{it})$ i $C_{it}^{\text{gr}} = C_{it}^{\text{syst}} EK_i = C_{it}^{\text{obs}} \exp(-v_{it}) EK_i$; $RK1_{it} = C_{it}^{\text{obs}} - C_{it}^{\text{syst}}$ oznacza różnicę kosztu obserwowanego i systematycznego, określającą zmianę kosztu wynikającą z zakłóceń losowych, obliczaną dla wartości uwzględniających nieefektywność;

$RK2_{it} = C_{it}^{\min} - C_{it}^{\text{gr}} = RK1_{it} \cdot EK_i$ to różnica kosztu minimalnego i granicznego, określająca zmianę kosztu wynikającą z działania czynników losowych, ale obliczona dla kosztu nie zawierającego nieefektywności;

$RK3_{it} = C_{it}^{obs} - C_{it}^{min} = C_{it}^{obs}(1 - EK_i) = C_{it}^{nadm}$ to różnica kosztu obserwowanego i niezbędnego, określająca koszt nadwyżkowy i zawierająca efekt kosztowy zakłóceń losowych;

$RK4_{it} = C_{it}^{syst} - C_{it}^{gr} = C_{it}^{syst}(1 - EK_i)$ określa różnicę między kosztem systematycznym i granicznym, oznaczającą efekt kosztowy nieefektywności przy pominięciu zakłóceń losowych.

Zauważmy na koniec tej części pracy, że dzięki wprowadzeniu nowych kategorii modelowych kosztu indywidualne wskaźniki efektywności kosztowej EK_i oraz wskaźniki RF_{it} , określające względną zmianę kosztu na skutek działania czynników losowych, mogą zostać wyznaczone na równoważne sposoby, odpowiednio:

$$EK_{it} = \frac{C_{it}^{min}}{C_{it}^{obs}} = \frac{C_{it}^{gr}}{C_{it}^{syst}} = \exp(-u_i), \quad RF_{it} = \frac{C_{it}^{obs}}{C_{it}^{syst}} = \frac{C_{it}^{min}}{C_{it}^{gr}} = \exp(v_{it}).$$

4. BAYESOWSKIE GRANICZNE MODELE KOSZTU DLA DANYCH PANELOWYCH

Podstawowym narzędziem przyjętym do uzyskania ocen efektywności kosztowej przedsiębiorstw dystrybucji energii jest bayesowski graniczny model kosztu stosowany dla danych przekrojowo-czasowych. Należy podkreślić, że wykorzystanie danych panelowych do wnioskowania o wskaźnikach efektywności poszczególnych obiektów pozwala na uwzględnienie znacznie szerszego zbioru zmiennych objaśniających i precyzyjniejszy szacunek niż w przypadku danych przekrojowych. Przy tak małej liczbie obiektów, jak w sektorze dystrybucji energii w Polsce po roku 2000, trudno liczyć na uzyskanie wiarygodnych wyników na podstawie danych z jednego roku lub danych uzyskanych przez uśrednienie obserwacji po czasie.

Bayesowskie graniczne funkcje kosztu dla danych panelowych zaproponowali Koop, Osiewalski i Steel (1994, 1997) (zob. też Fernández, Osiewalski i Steel, 1997), definiując m.in. model z losowymi efektami indywidualnymi o wspólnym rozkładzie efektywności (ang. *Common Efficiency Distribution* — CED), wykorzystany w naszych badaniach empirycznych.

Bayesowski graniczny model kosztu, w przypadku nieefektywności traktowanej jako efekt w indywidualny i przy przyjęciu specyfikacji CED, określony jest przez łączny rozkład macierzy obserwacji $Y = [y_{it} \ (i = 1, \dots, N; \ t = 1, \dots, T)]$, wektora zmiennych ukrytych $u = [u_1 \ \dots \ u_N]'$, $k + 1$ elementów wektora β , precyzji σ_v^{-2} symetrycznego składnika losowego i parametru φ rozkładu nieefektywności, przy ustalonej macierzy X zmiennych egzogenicznych. Model ten zapisujemy w postaci:

$$p(Y, \beta, \sigma_v^2, u, \varphi | X) = p(\beta, \sigma_v^2, \varphi) \prod_{i=1}^N \left[f_G(u_i | 1, \varphi) \prod_{t=1}^T f_N(y_{it} | x_{it} \beta + u_i, \sigma_v^2) \right] \quad (1)$$

gdzie $f_N(\cdot|a, b)$ i $f_G(\cdot|c, d)$ oznaczają odpowiednio funkcje gęstości: rozkładu normalnego o wartości oczekiwanej a i wariancji b oraz rozkładu gamma o wartości oczekiwanej c/d^2 i wariancji c/d^2 ; $p(\beta, \sigma_v^{-2}, \varphi) = p(\beta)p(\sigma_v^{-2})p(\varphi)$ jest łącznym rozkładem a priori dla β , σ_v^{-2} oraz φ . Zauważmy, że warunkowy względem parametrów rozkład gamma dla u_i to rozkład wykładniczy o wartości oczekiwanej i odchyleniu standardowym $1/\varphi$.

Łączna gęstość a priori jest zdefiniowana jako iloczyn gęstości brzegowych dla β , σ_v^{-2} oraz φ . Dla φ przyjęto rozkład gamma z $c = 1$ i $d = -\ln(r^*)$, czyli wykładniczy o wartości oczekiwanej $-1/\ln(r^*)$, gdzie stała r^* jest medianą a priori efektywności EK_i (van den Broeck, Koop, Osiewalski i Steel, 1994); ustalono $r^* = 0,8$. Dla precyzji symetrycznego składnika losowego σ_v^{-2} założono rozkład z rodziny gamma o gęstości $f_G(\sigma_v^{-2} | g_1/2, g_2/2)$, z $g_1 = NT-k-1$ oraz $g_2 = 10^{-4}$, natomiast dla wektora parametrów strukturalnych β możemy przyjąć rozkład jednostajny (ucięty przez warunki regularności ekonomicznej wynikające z własności funkcji kosztu zmiennego) bądź rozkład normalny o wektorze wartości oczekiwanych a priori β^0 i macierzy precyzji Σ . W naszych badaniach przyjmujemy bardzo rozproszony rozkład $N(0, 10I_{k+1})$, rezygnując z narzucania warunków regularności mikroekonomicznej. Nasz rozkład a priori odzwierciedla więc praktycznie brak wstępnej wiedzy o parametrach innych niż φ .

Łączny rozkład a posteriori dla parametrów i zmiennych ukrytych, uzyskany w modelu bayesowskim (1) i o gęstości proporcjonalnej do (1), jest niestandardowym rozkładem określonym na przestrzeni o wymiarze równym sumie liczby zmiennych ukrytych (czyli obiektów, N) i parametrów. Skomplikowana postać gęstości tego rozkładu nie pozwala na analityczne wyznaczenie momentów i rozkładów brzegowych dla żadnego z parametrów czy efektów indywidualnych występujących w modelu, natomiast umożliwia wyprowadzenie (dla ich ustalonych bloków) układu warunkowych rozkładów a posteriori. Te pełne rozkłady warunkowe są standardowe (gamma, normalne, ucięte normalne), umożliwiając generowanie liczb losowych według schematu Gibbsa, tj. metody z rodziny Monte Carlo łańcuchów Markowa (ang. *Markov Chain Monte Carlo* — MCMC), a w konsekwencji uzyskiwanie prób z łącznego rozkładu a posteriori i łatwe przybliżanie jego charakterystyk. Koop, Steel i Osiewalski (1995) po raz pierwszy pokazali układ pełnych warunkowych rozkładów a posteriori i zastosowali losowanie Gibbsa w bayesowskich modelach granicznych, ukazując przewagę tego podejścia w stosunku do metody Monte Carlo z funkcją ważności (ang. *Monte Carlo Importance Sampling* — MCIS), którą stosowali van den Broeck, Koop, Osiewalski i Steel (1994) w pierwszej pracy z zakresu bayesowskiej analizy efektywności; zob. też Osiewalski i Steel (1998). Podstawy losowania Gibbsa i stosowny kod komputerowy (w języku pakietu GAUSS) opracowali dla modeli z nieujemnymi efektami indywidualnymi (dla danych panelowych) Koop, Osiewalski i Steel (1997); było to podstawą wielu dalszych zastosowań (por. Marzec i Osiewalski, 2003, 2008; Wróbel-Rotter

i Osiewalski, 2002; Osiewalski, 2001), w tym badań empirycznych tej pracy. Teoretyczne uzasadnienie i własności metod MCMC, w tym losowania Gibbsa, prezentuje Tierney (1994); zob. też np. O'Hagan (1994).

5. WYNIKI BADAŃ EMPIRYCZNYCH

Analiza empiryczna została przeprowadzona na danych rocznych pochodzących z czternastu przedsiębiorstw dystrybuujących energię elektryczną (oznaczonych SD01–SD14), obserwowanych w okresie sześciu lat (2001–2006). Podstawowym narzędziem analizy efektywności kosztowej jest bayesowski model o wspólnym rozkładzie efektywności (CED), w którym koszt zmienny opisany jest funkcją Cobba i Douglasa z argumentami oddającymi skalę usług (produkcji) i sieć dystrybucyjną rozważanych podmiotów.

Modelowany koszt zmienny to koszt operacyjny (suma kosztu obrotu i dystrybucji), który obejmuje głównie wynagrodzenia wraz z narzutami oraz wydatki na usługi obce i materiały. Zmienne objaśniające i różne warianty modelu kosztu zostały wyspecyfikowane w konsultacji z ekspertami z Urzędu Regulacji Energetyki (URE), który dostarczył danych.

5.1. Dobór zmiennych objaśniających a pomiar efektywności kosztowej

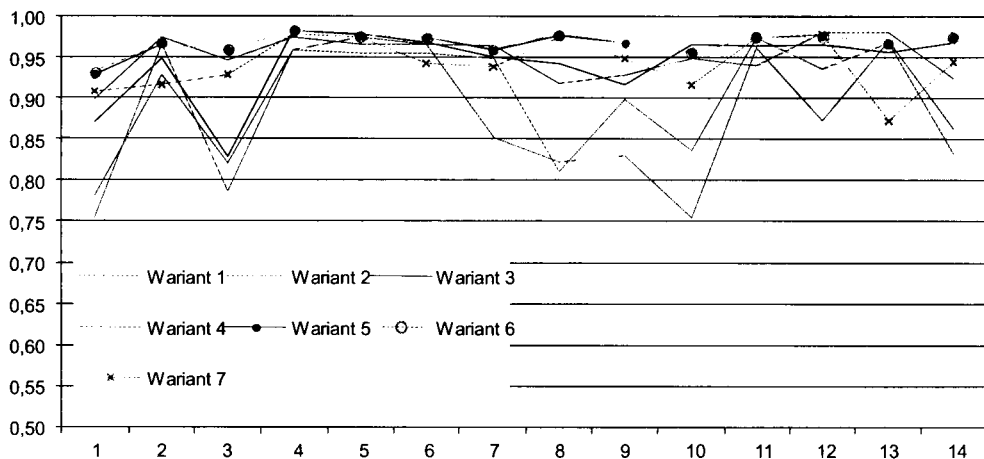
Zbiór czynników wyjaśniających zawiera takie zmienne jak: długość przesyłowych linii napowietrznych i kablowych wysokiego, średniego i niskiego napięcia (WN, SN i nN ujętych w różnych kategoriach), liczba odbiorców i dostawy energii w poszczególnych grupach, zmienne techniczne opisujące sieć: liczba i moc transformatorów oraz liczba stacji elektroenergetycznych. Wstępna lista zmiennych, zaczerpnięta z opracowania URE „Taryfy Spółek Dystrybucyjnych na okres 2002/2003”, posłużyła do budowy konkurencyjnych wariantów granicznej funkcji kosztu, mających na celu ilustrację wrażliwości wnioskowania o wskaźnikach efektywności i parametrach strukturalnych, a w efekcie określenie zmiennych, które mają największy wpływ na kształtowanie się kosztów operacyjnych. Liczba zmiennych w zależności od wariantu modelu waha się w granicach od 11 do 22, zob. tabela 1.

Bayesowskie oceny efektywności obiektów, uzyskane w różnych wariantach modelu kosztu, a także średnie tych ocen, przedstawiono na rycinach 1 i 2. Średni poziom efektywności kosztowej sektora zależy od zbioru zmiennych objaśniających i jest najwyższy w przypadku modeli o dwudziestu i więcej zmiennych; dla tych modeli różnice ocen między obiektami najmniej i najbardziej efektywnym są najmniejsze. Oceny efektywności obiektów 4, 5, 6 i 11 są niewrażliwe na dobór zmiennych, bardzo wrażliwe są oceny obiektów 1, 3, 8

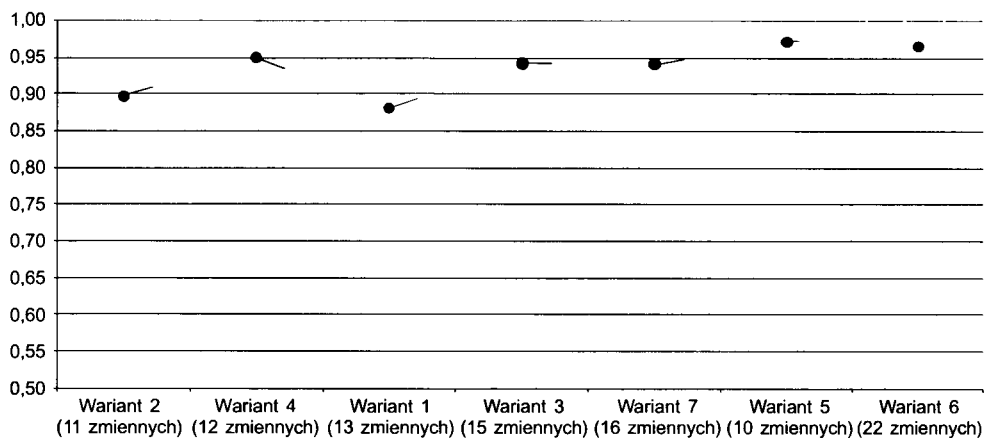
Lista zmiennych objaśniających rozważanych w modelowaniu kosztu operacyjnego

		Wariant 7	Wariant 6	Wariant 5	Wariant 4	Wariant 3	Wariant 2	Wariant 1
1	Całkowita długość linii WN w przeliczeniu na jeden tor linii	km	x	x	x	x	x	x
2	Całkowita długość linii SN w przeliczeniu na jeden tor linii	km			x	x	x	
3	Całkowita długość linii nN w przeliczeniu na jeden tor linii	km			x		x	
4	Długość linii SN napowietrznych w przeliczeniu na jeden tor linii	km	x	x	x			x
5	Długość linii SN kablowych w przeliczeniu na jeden tor linii	km	x	x	x			x
6	Długość linii nN napowietrznych w przeliczeniu na jeden tor linii + długość przyłączy napowietrznych nN	km	x	x	x	x		
7	Długość linii nN kablowych w przeliczeniu na jeden tor linii + długość przyłączy kablowych nN	km	x	x	x	x		
8	Całkowita długość linii nN w przeliczeniu na jeden tor linii + długość przyłączy nN	km						x
9	Całkowita moc transformatorów	MVA			x	x		x
10	Liczba stacji elektroenergetycznych SN i nN	szt.	x	x	x		x	
11	Liczba stacji elektroenergetycznych 110 kV	szt.	x	x	x		x	
12	Moc transformatorów WN/SN	MVA	x	x	x		x	

13	Moc transformatorów SN/nN	MVA	x	x	x			x	
14	Całkowita liczba transformatorów	szt.	x	x	x				x
15	Całkowita liczba stacji elektroenergetycznych	szt.				x	x		x
16	Liczba miejsc dostarczania u odbiorców na WN	szt.	x	x	x	x	x		
17	Liczba odbiorców na SN	szt.	x	x	x	x	x		
18	Liczba odbiorców na nN	szt.				x			
19	Liczba odbiorców na nN w gr. tar. C	szt.	x	x	x		x	x	x
20	Liczba odbiorców na nN w gr. tar. G	szt.	x	x	x		x	x	x
21	Dostawa energii odbiorcom na SN	MWh		x	x	x	x		
22	Dostawa energii na nN	MWh				x			
23	Dostawa energii na nN w gr. tar. C	MWh		x	x		x	x	x
24	Dostawa energii na nN w gr. tar. G	MWh		x	x		x	x	x
25	Dostawa energii odbiorcom na WN	MWh		x	x	x	x		
26	Średnia moc szczytowa netto	MW		x					x
27	Energia pobrana w transzycie ogółem (WN, SN, nN) + Energia wprowadzona z sieci PSE S.A.	MWh		x	x				x
28	Dostawa energii ogółem	MWh	x						
29	Przeciętne zatrudnienie	osoby		x		x			
30	Zmienna czasowa <i>t</i>	lata	x	x	x		x		
Liczba zmiennych w modelu			16	22	20	12	15	11	13



Ryc. 1. Wartości oczekiwane *a posteriori* indywidualnych wskaźników efektywności dla konkurencyjnych wariantów modelu



Ryc. 2. Wartości oczekiwane *a posteriori* średniego wskaźnika efektywności spółek dla konkurencyjnych wariantów modelu w zależności od liczby zmiennych niezależnych

i 10. Wrażliwość ta wynika z uwzględniania w rozbudowanych wariantach modelu zmiennych objaśniających szczególnie ważnych dla niektórych przedsiębiorstw, np. brak w wariantach 1 i 2 zarówno dostawy odbiorcom na WN, jak i liczby miejsc dostarczania, będących szczególnie istotnymi charakterystykami profilu produkcji niektórych spółek, prowadzi do obniżenia wartości tej części ich kosztu, która ma ekonomiczne uzasadnienie w ramach modelu.

Średnia efektywność kosztowa waha się (w zależności od przyjętego zestawu zmiennych egzogenicznych) od 88% ($0,882 \pm 0,052$) do 97% ($0,970 \pm 0,027$), co oznacza, że średnio tylko kilka procent kosztu obserwowanego było kosztem nieuzasadnionym.

Należy pamiętać, że jeśli modelujemy koszt zmienny, obejmujący wynagrodzenia, to powinien być on objaśniany przez wielkości produkcji, ceny czynników zmiennych i nakłady czynników stałych, ale nie przez (zawarty w koszcie) nakład czynnika zmiennego, jakim jest zatrudnienie. Obecność tej zmiennej w wariantach 4 i 6 czyni je trudno interpretowalnymi w świetle teorii kosztu zmiennego¹. Zauważmy, że żaden wariant nie zawiera cen czynników zmiennych (zwłaszcza płac), a zmienne objaśniające (poza zatrudnieniem i zmienną czasową) reprezentują produkcję i czynniki stałe — elementy kapitału rzeczowego. Zleceniodawca nie chciał uwzględniać cen czynników produkcji, co można interpretować jako żądanie analizy kosztu w hipotetycznych warunkach jednakowych cen czynników (płac) dla wszystkich przedsiębiorstw, gdyż regulator nie jest zainteresowany uzasadnianiem wysokich kosztów wysokimi płacami. Przy założeniu funkcji Cobba i Douglasa oznacza to, że efekt cen (stałych po obiektach) ukryty jest w wyrazie wolnym. Ponieważ trudno jest traktować ceny czynników zmiennych (płace) jako stałe w czasie, wyraz wolny powinien być uzupełniony o zmienną czasową, reprezentującą ewentualne trendy w cenach; zmienna ta może również służyć (w sposób uproszczony, lecz uzasadniony krótkim szeregiem czasowym) uwzględnieniu efektów postępu technologicznego. Wpływ (na koszt) wzrostu cen jest dodatni (koszt wzrasta ze wzrostem ceny), natomiast wpływ postępu technologicznego jest ujemny (postęp obniża koszt produkcji). Wariant 5, spełniający postulaty poprawnej ekonomicznie specyfikacji kosztu zmiennego oraz adekwatny ze statystycznego punktu widzenia, stanowi postać ostateczną.

5.2. Wnioskowanie o parametrach funkcji kosztu zmiennego

Tabela 2 przedstawia wartości oczekiwane i odchylenia standardowe *a posteriori* parametrów funkcji kosztu Cobba i Douglasa w wariancie 5. Są one interpretowane jako elastyczności kosztu względem danego czynnika wyjaśniającego i informują, o ile procent byłby wyższy koszt operacyjny, gdyby wybrana zmienna egzogeniczna wzrosła o jeden procent, przy ustalonych wartościach pozostałych czynników. Na uwagę zasługują oceny parametrów przy 4 zmiennych (o numerach z zakresu 21–25) określających dostawy energii, które należy uznać za produkty spółek dystrybucyjnych. Oceny parametrów przy nich są podobne w alternatywnych wariantach modelu, przy czym współczynniki

¹ Przy braku zmiennej płacowej można je uważać za pewne przybliżenia modelu kosztu całkowitego z pracą jako czynnikiem stałym, w którym jednak zbyt wysokie zatrudnienie nie znajdzie odzwierciedlenia w niskiej efektywności. Regulator nie powinien zatem być zainteresowany traktowaniem pracy jako czynnika stałego, gdyż może to zawyżać ocenę efektywności.

Wartości oczekiwane i odchylenia standardowe a posteriori parametrów funkcji kosztu (wariant 5)

Wstępna lista zmiennych		$E(.)$	$D(.)$	\sqrt{NVIF}
0	Wyraz wolny	-1,2489	0,9762	464.9
1	Całkowita długość linii WN w przeliczeniu na jeden tor linii	0,3827*	0,0924	249.7
2	Całkowita długość linii SN w przeliczeniu na jeden tor linii	—	—	
3	Całkowita długość linii nN w przeliczeniu na jeden tor linii	—	—	
4	Długość linii SN napowietrznych w przeliczeniu na jeden tor linii	-0,2063*	0,0600	252.5
5	Długość linii SN kablowych w przeliczeniu na jeden tor linii	-0,0642	0,1029	196.4
6	Długość linii nN napowietrznych w przeliczeniu na jeden tor linii + długość przyłączy napowietrznych nN	0,3643*	0,0630	160.3
7	Długość linii nN kablowych w przeliczeniu na jeden tor linii + długość przyłączy kablowych nN	0,2921*	0,0697	174.9
8	Całkowita długość linii nN w przeliczeniu na jeden tor linii + długość przyłączy nN	—	—	
9	Całkowita moc transformatorów	—	—	
10	Liczba stacji elektroenergetycznych SN i nN	-0,5769*	0,1649	816.0
11	Liczba stacji elektroenergetycznych 110 kV	-0,1441	0,0941	128.2
12	Moc transformatorów WN/SN	-0,1166	0,0990	223.9
13	Moc transformatorów SN/nN	-0,5473*	0,1716	431.6
14	Całkowita liczba transformatorów	0,6050*	0,1161	868.2
15	Całkowita liczba stacji elektroenergetycznych	—	—	
16	Liczba miejsc dostarczania u odbiorców na WN	-0,0148*	0,0064	10.1
17	Liczba odbiorców na SN	0,3453*	0,0603	158.8
18	Liczba odbiorców na nN	—	—	
19	Liczba odbiorców na nN w gr. tar. C	-0,4539*	0,1502	752.2
20	Liczba odbiorców na nN w gr. tar. G	-0,9559*	0,1989	766.3
21	Dostawa energii odbiorcom na SN	-0,1213	0,0625	215.1
22	Dostawa energii na nN	—	—	
23	Dostawa energii na nN w gr. tar. C	1,1919*	0,1351	553.6
24	Dostawa energii na nN w gr. tar. G	0,9436*	0,0974	455.6
25	Dostawa energii odbiorcom na WN	0,0125	0,0083	60.1
26	Średnia moc szczytowa netto	—	—	

Wstępna lista zmiennych		$E(.)$	$D(.)$	\sqrt{NVIF}
27	Energia pobrana w tranzycie ogółem (WN, SN, nN) + Energia wprowadzona z sieci PSE S.A.	-0,0305*	0,0053	67.7
28	Dostawa energii ogółem	—	—	
29	Przeciętne zatrudnienie	—	—	
30	Zmienna czasowa t	0,0035	0,0030	3.9

$E(.)$ — wartość oczekiwana *a posteriori*, $D(.)$ — odchylenie standardowe *a posteriori*.

Objaśnienia: * wartość bezwzględna wartości oczekiwanej jest co najmniej dwukrotnie większa niż odchylenie standardowe.

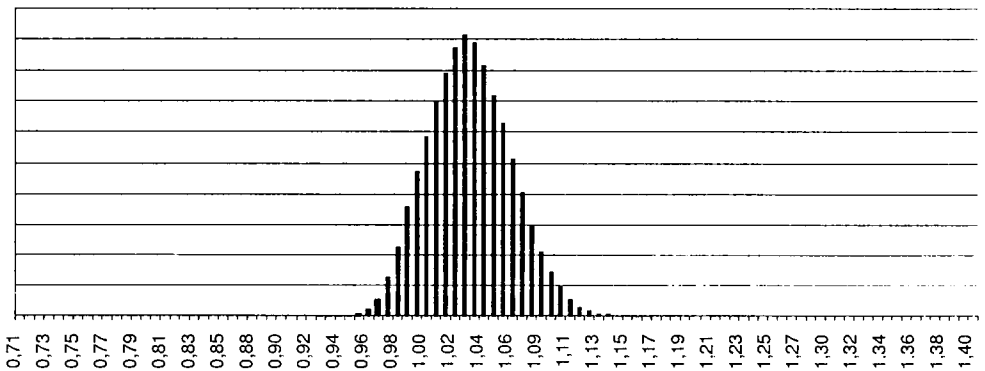
bardzo precyzyjnie szacowane notuje się dla zmiennych charakteryzujących dostawę energii na niskim napięciu (nN), oznaczające wzrost kosztu operacyjnego o około 2% na skutek wzrostu dostawy energii nN o 1%, *ceteris paribus*. Zmienna określająca dostawę energii na wysokim napięciu (WN) ma niewielki, dodatni wpływ na koszt operacyjny (rzędu 0,0125% w odpowiedzi na wzrost dostawy o 1%), zaś wpływ zmiennej opisującej dostawę energii na średnim napięciu (SN) ma znak ujemny we wszystkich przypadkach; parametry te są nieprecyzyjnie szacowane, co może wskazywać na nieistotność wpływu tych zmiennych na koszt operacyjny.

Analogiczną interpretację można przeprowadzić dla każdej z pozostałych zmiennych (z wyjątkiem czasowej, nieistotnej), np. łącznemu wzrostowi liczby odbiorców na nN o 1% odpowiada więcej niż proporcjonalna obniżka kosztu operacyjnego (o ok. 1,4%), wzrostowi liczby odbiorców WN o 1% odpowiada spadek kosztu o ok. 0,015%, zaś zwiększenie liczby odbiorców energii na SN jest związane ze zwiększeniem kosztu o ok. 0,35%. Zwiększenie liczby stacji elektroenergetycznych i mocy transformatorów ma ujemny wpływ na koszt operacyjny, natomiast liczba transformatorów — dodatni. Wszystkie 14 zmiennych (z zakresu 1–20) charakteryzują kapitał rzeczowy przedsiębiorstw dystrybucji energii w różnych jego kategoriach i aspektach. Równoczesne zwiększenie tych 14 zmiennych o 1% jest związane ze zmianą kosztu operacyjnego o -1,09% ($\pm 0,3\%$). Zmienną nr 27 traktujemy też jako nakład czynnika stałego (energia na wejściu systemu); jego wzrost o 1% jest związany ze spadkiem kosztu o ok. 0,03%.

Zauważmy, że suma elastyczności kosztu względem produktów jest odwrotnością (radialnego) współczynnika efektu skali (por. Varian, 1992). W naszym przypadku, w którym badamy koszt zmienny, wartość oczekiwana *a posteriori* sumy tych 4 elastyczności wynosi 2,03 (odchylenie standardowe 0,25), a zatem mamy niski (równy ok. 0,5) poziom malejącego krótkookresowego efektu skali. Przybliżonej oceny całkowitego (długookresowego) efektu skali dokonamy na podstawie formuły:

$$RTS = \frac{1 - \sum \partial \ln VC / \partial \ln K_i}{\sum \partial \ln VC / \partial \ln Q_j},$$

gdzie K jest symbolem nakładu czynnika stałego a Q — wielkości produkcji; zob. Mąrzec i Osiewalski (2008). Wstawiając w miejsce elastyczności ich wartości oczekiwane *a posteriori* z tabeli 2, otrzymujemy $RTS = [1 - (-1,09 - 0,03)] / 2,03 = 1,05$, czyli ocenę wskazującą na stały bądź niewielki rosnący całkowity efekt skali. Histogram na rycinie 3 ukazuje brzegowy rozkład *a posteriori* dla RTS (uzyskany w symulacji MCMC), którego wartość oczekiwana wynosi 1,03 a odchylenie standardowe 0,03. Wnioskowanie o całkowitym efekcie skali jest stosunkowo precyzyjne (wziąwszy pod uwagę, że jest wyznaczany z krótkookresowej funkcji kosztu zmiennego); nie można odrzucić hipotezy, że jest on stały. Jest to bardzo ciekawy wynik modelowania technologii sektora dystrybucji energii w Polsce.



Ryc. 3. Rozkład *a posteriori* długookresowego efektu skali

Należy pamiętać, że obecność w modelu tak wielu merytorycznie powiązanych ze sobą zmiennych objaśniających może skutkować ich (przybliżoną) współliniowością, która prowadzi — przy mało informacyjnym rozkładzie *a priori* parametrów funkcji kosztu — do niskiej precyzji szacunku indywidualnych parametrów. Niebayesowskie mierniki wpływu braku ortogonalności kolumn macierzy X na precyzję szacunku omawia Osiewalski (1992); są to tzw. niescentrowane współczynniki zwiększenia wariancji (*NVIF*) estymatora MNK. Odpowiadają one miernikom bayesowskim dla wariancji warunkowego rozkładu *a posteriori* wektora β przy braku informacji *a priori*. W interpretacji bayesowskiej, $NVIF(c'\beta)$ informuje, ile razy warunkowa wariancja *a posteriori* liniowej kombinacji parametrów regresji liniowej jest większa przy danej macierzy X w stosunku do hipotetycznej macierzy o ortogonalnych kolumnach (ale tej samej długości, co kolumny oryginalne). Podstawą pomiaru odchylenia macie-

rzy X od ortogonalności oraz analizy własności miernika $NVIF$ są niescentrowane współczynniki korelacji, tworzące macierz $R_N = (XW^{-1})'(XW^{-1}) = W^{-1}(X'X)W^{-1}$, gdzie W jest macierzą diagonalną z długościami kolumn macierzy X na przekątnej. W omawianej funkcji kosztu macierz X ma 21 kolumn, więc macierz R_N ma 21 wartości własnych, z których największa to 20,64, zaś najmniejsza to zaledwie $6,377 \cdot 10^{-7}$. Odchylenia od ortogonalności są więc silne; najmniejszy niescentrowany współczynnik korelacji wynosi 0,8162, znaczna ich większość przekracza 0,99. Wzrost warunkowego odchylenia standardowego dla $c'\beta$ na skutek braku ortogonalności zależy od c ; może wynosić od $(20,64)^{-0,5} = 0,22$ raza (zwiększenie precyzji szacunku) do aż $(6,377 \cdot 10^{-7})^{-0,5} = 1252$ razy (ogromny spadek precyzji szacunku). Iloraz pierwiastków kwadratowych największej i najmniejszej wartości własnej, czyli u nas $1252/0,22 = 5689$, pokrywa się z indeksem uwarunkowania macierzy X ; zob. Belsley i in., 1980. Przypomnijmy, że dla indywidualnego współczynnika regresji wzrost precyzji szacunku (na skutek braku ortogonalności) nie jest możliwy, gdyż jego $NVIF$ jest równy odpowiedniemu elementowi z przekątnej macierzy $(R_N)^{-1}$, czyli liczbie nie mniejszej od 1; pierwiastek kwadratowy z $NVIF$ pokazano w ostatniej kolumnie tabeli 2. Na tle stosunkowo silnych zależności między 21 zmiennymi objaśniającymi, uzyskane rozkłady *a posteriori* parametrów funkcji kosztu i ich transformacji wydają się wystarczająco precyzyjne dla potrzeb interpretacji.

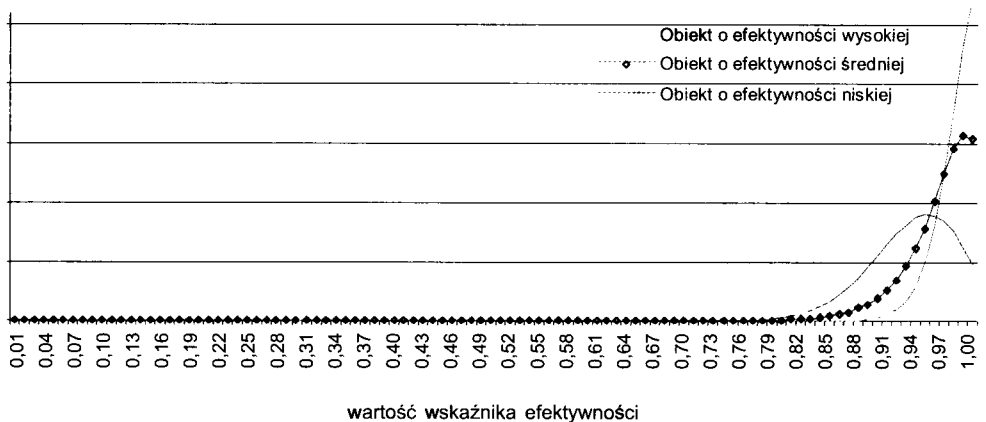
5.3. Wnioskowanie o efektywności i kategoriach modelowych kosztu

Tabela 3 przedstawia uzyskane w wariacie 5 charakterystyki rozkładu *a posteriori* wskaźników efektywności (wartości oczekiwane, odchylenia standardowe) i ranking spółek według malejącej wartości oczekiwanej. Rycina 4 prezentuje brzegowe gęstości *a posteriori* efektywności trzech firm, zajmujących w rankingu miejsca: drugie (SD05), dziewiąte (SD02) i ostatnie (SD03). Wyniki wskazują na dość wysoką precyzję wnioskowania.

Wnioskowanie bayesowskie na podstawie danych przekrojowo-czasowych opiera się na $NT = 84$ obserwacjach ($T = 6$, $N = 14$), co zapewnia stabilność wyników, wykorzystanie całej dostępnej informacji o analizowanych obiektach i umożliwi rozpatrzenie większego zbioru zmiennych egzogenicznych niż w przypadku estymacji na danych przekrojowych ($N = 14$). Ma to istotne znaczenie w przypadku estymacji indywidualnej efektywności, która w przypadku danych przekrojowo-czasowych jest szacowana dla każdego z obiektów na podstawie $T = 6$ obserwacji, natomiast w przypadku danych przekrojowych jedynie na podstawie $T = 1$. Model dla danych przekrojowo-czasowych, wykorzystujący bezpośrednio zmienne egzogeniczne wpływające na poziom kosztu (a nie zmienne agregatowe, stosowane poprzednio przez Regulatora), umożliwi

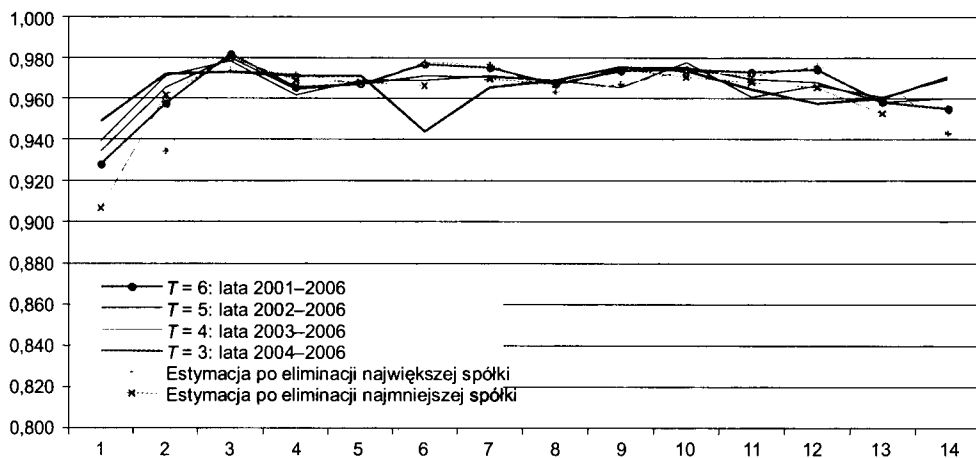
Wyniki w zakresie indywidualnej efektywności kosztowej

Spółka	Wartości oczekiwane <i>a posteriori</i>	Miejsce	Odchylenie standardowe <i>a posteriori</i>
SD01	0,923	13	0,052
SD02	0,954	9	0,034
SD03	0,865	14	0,068
SD04	0,980	1	0,017
SD05	0,979	2	0,019
SD06	0,949	11–12	0,037
SD07	0,949	11–12	0,045
SD08	0,971	4–5	0,025
SD09	0,952	10	0,042
SD10	0,958	8	0,032
SD11	0,971	4–5	0,024
SD12	0,973	3	0,022
SD13	0,965	6–7	0,030
SD14	0,965	6–7	0,026
Średnio	0,954		0,034

Ryc. 4. Rozkład *a posteriori* wskaźników efektywności kosztowej dla wybranych obiektów

wia więc nie tylko określenie ich indywidualnego wpływu na koszt obserwowany, ale również precyzyjną ocenę wskaźników efektywności kosztowej.

Estymację modelu (w wariancie 5) powtórzono na zbiorze danych uzyskanych po eliminacji największej i najmniejszej spółki oraz na zbiorach powstałych po skróceniu szeregu czasowego (np. dla $T = 5$ model oszacowano na podstawie danych od roku 2002 do 2006, dla $T = 4$ od 2003 do 2006 itd., ryc. 5).

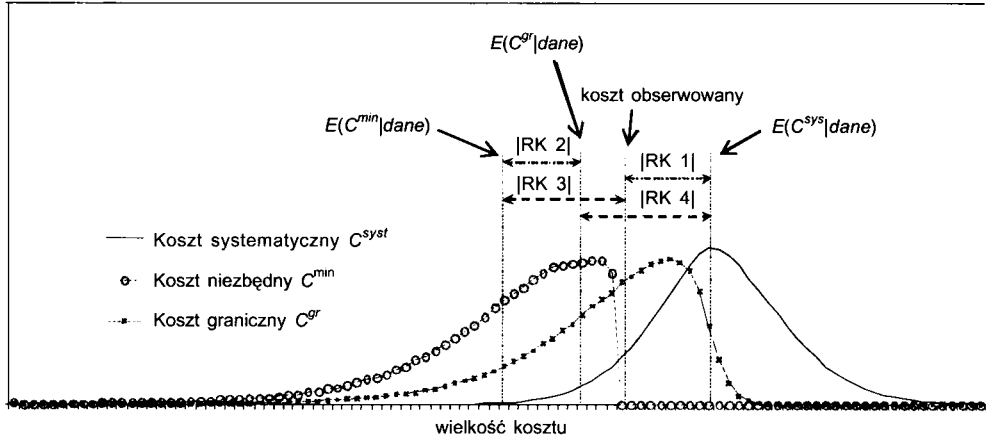


Ryc. 5. Oceny indywidualnych wskaźników efektywności dla modeli szacowanych po ograniczeniu liczby obserwacji w czasie

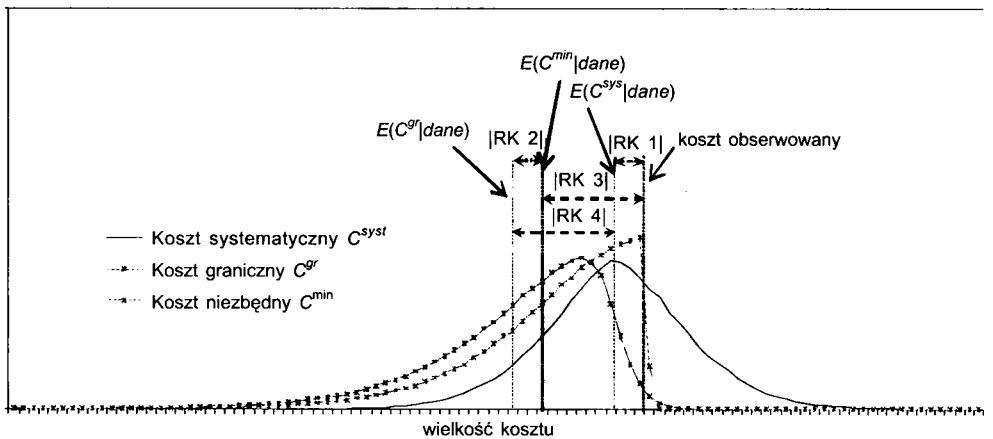
Otrzymane oceny wskaźników efektywności są podobne; świadczy to o stabilności i odporności wyników oraz potwierdza pośrednio poprawność modelu. Ranking spółek nie ulega zasadniczym zmianom; obserwuje się jedynie niewielkie przemieszczenie spółek przy zachowaniu miejsca w grupie efektywniejszych bądź mniej efektywnych obiektów. Do poszczególnych miejsc rankingu nie należy przywiązywać znaczenia, gdyż różnice wartości oczekiwanych *a posteriori* indywidualnych efektywności są niewielkie w stosunku do odchyleń standardowych. Prezentowane wyniki są niewrażliwe na zmiany (w przedziale 0,5–0,95) wartości r^* , tj. kluczowej stałej rozkładu *a priori*.

Obserwowany koszt operacyjny można poddać dekompozycji, wykorzystując koszty teoretyczne wprowadzone w części 3 (graniczny, niezbędny, systematyczny, nadwyżkowy). W roku 2006 średni obserwowany koszt wynosił 308 400 tys. zł, średni koszt systematyczny (zawierający średnią nieefektywność przedsiębiorstw) oszacowany na podstawie modelu, po usunięciu wpływu zakłóceń losowych oraz czynników nie ujętych w modelu, jest równy 303 185 tys. zł (± 6501 tys. zł). Dla części obiektów koszt systematyczny jest wyższy od obserwowanego; w ich przypadku czynniki losowe powodują, że koszt rze-

czywisty jest niższy od kosztu, jaki teoretycznie spółki te powinny ponosić przy danym poziomie działalności i indywidualnej nieefektywności. W pozostałych przypadkach czynniki losowe powodują zwiększenie kosztu obserwowanego w stosunku do systematycznego. Średni poziom kosztu niezbędnego, zawierającego efekt czynników losowych (ale bez nieefektywności), został dla roku 2006 oszacowany na poziomie 296 324 tys. zł (± 9086 tys. zł). Ryciny 6 i 7 prezentują (dla dwóch spółek) brzegowe gęstości *a posteriori* teoretycznych kategorii kosztu; ukazują one, jak nieprecyzyjne (obarczone niepewnością) jest wnioskowanie o tych nieobserwowalnych wielkościach.

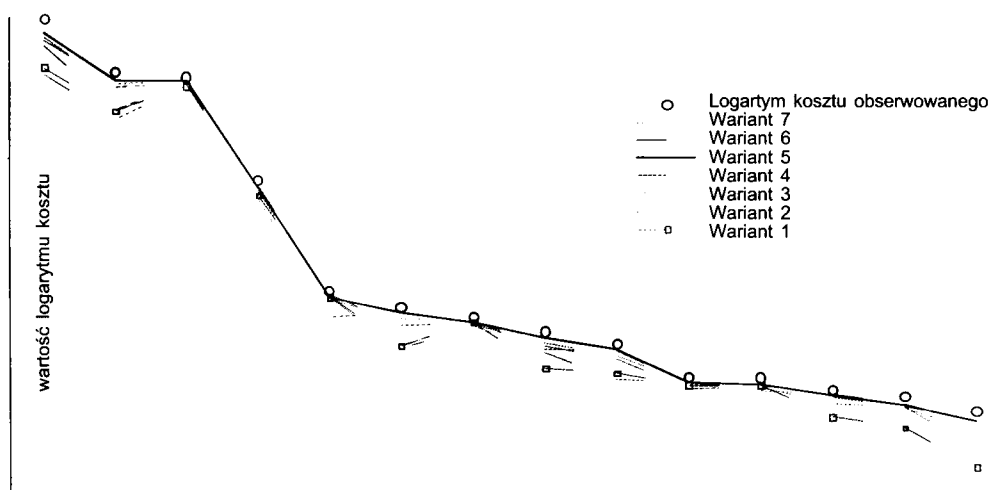


Ryc. 6. Rozkłady *a posteriori* kosztu granicznego, niezbędnego i systematycznego dla SD02 w roku 2006



Ryc. 7. Rozkłady *a posteriori* kosztu granicznego, niezbędnego i systematycznego dla SD06 w roku 2006

Konstrukcja modelu granicznego zapewnia, że przy wyznaczaniu kosztu minimalnego brana jest pod uwagę specyfika funkcjonowania poszczególnych przedsiębiorstw, która może mieć dodatni albo ujemny wpływ na koszt obserwowany. Koszt nadwyżkowy, który powinien zostać zredukowany, jest zdefiniowany jako różnica między kosztem rzeczywistym a kosztem niezbędnym zawierającym efekt wpływu czynników losowych. Ocena kosztu niezbędnego dla konkurencyjnych zestawów zmiennych egzogenicznych (ryc. 8) wykazuje różnice: średni koszt niezbędny dla 2006 waha się w przedziale od 268 593 tys. zł do 296 324 tys. zł, w sytuacji kiedy koszt obserwowany wynosi 308 400 tys. zł. Najbliższy obserwowanemu jest koszt niezbędny w wariancie 5, w którym otrzymano wysokie oceny efektywności kosztowej spółek.



Ryc. 8. Porównanie kosztu obserwowanego w 2006 roku i wartości oczekiwanych *a posteriori* kosztu minimalnego w konkurencyjnych wariantach modelu

6. PODSUMOWANIE

Ocena efektywności działania 14 spółek dystrybucyjnych zaprezentowana w pracy opiera się na modelu ekonometrycznym, opisującym kształtowanie się kosztu operacyjnego w zależności od szeregu czynników o charakterze techniczno-ekonomicznym. Zbiór potencjalnych zmiennych objaśniających zawiera 30 wielkości, które posłużyły do estymacji modelu w konkurencyjnych wariantach, mających na celu analizę wrażliwości wnioskowania o parametrach strukturalnych i wskaźnikach efektywności oraz określenie wpływu poszczególnych zmiennych na poziom kosztu operacyjnego. Uzyskane wyniki są po-

mocne w określeniu (dla każdego z przedsiębiorstw) wartości uzasadnionego kosztu operacyjnego przy danym poziomie działalności.

Zaprezentowany model ekonometryczny dla danych przekrojowo-czasowych pozwala na uwzględnienie podczas estymacji znacznie większej liczby zmiennych niż w przypadku analizy na danych przekrojowych, nie wymagając budowy zmiennych agregatowych, co w konsekwencji nie powoduje utraty dostępnej informacji oraz zapewnia uzyskanie stabilnych wyników. Metody estymacji bayesowskiej pozwalają na estymację dla każdego z obiektów (na podstawie obserwacji z kilku lat) wskaźników efektywności traktowanych jako efekty indywidualne (stałe w czasie), co prowadzi do wiarygodnych i interpretowalnych ekonomicznie rezultatów.

Stochastyczny model graniczny przyjęty w pracy umożliwia precyzyjną dekompozycję kosztu obserwowanego na koszt niezbędny i koszt nadwyżkowy. Koszt niezbędny uwzględnia efekt wynikający z zakłóceń losowych i innych czynników nieuwzględnionych w modelu oraz podstawowy mikroekonomiczny koszt graniczny, stanowiący wielkość uzasadnioną z teoretyczno-ekonomicznego punktu widzenia w celu zapewnienia danego poziomu działalności. Różnica między kosztem obserwowanym a kosztem niezbędnym określa koszt nadwyżkowy, który powinien zostać zredukowany. Indywidualny wskaźnik efektywności jest miarą procentową stopnia uzasadnienia poniesionego kosztu. Należy podkreślić, że model graniczny szacowany technikami bayesowskimi umożliwia obliczenie miar niepewności związanych z estymacją każdej z kategorii kosztów i efektywności, co ma istotne znaczenie praktyczne, gdyż informuje o precyzji wnioskowania.

BIBLIOGRAFIA

- Aigner D., Lovell C.A.K., Schmidt P. 1977. *Formulation and estimation of stochastic frontier production function models*, Journal of Econometrics, 6.
- Belsley D.A., Kuh E., Welsh R.E. 1980. *Regression Diagnostics*, J. Wiley, New York.
- Broeck van den J., Koop G., Osiewalski J., Steel M.F.J. 1994. *Stochastic frontier models: A Bayesian perspective*, Journal of Econometrics, 61.
- Fernández C., Osiewalski J., Steel M.F.J. 1997. *On the use of panel data in stochastic frontier models with improper priors*, Journal of Econometrics, 79.
- Greene W.H. 1993. *The econometric approach to efficiency analysis*, [w:] *The Measurement of Productive Efficiency — Techniques and Applications*, red.: H.O. Fried, C.A.K. Lovell, P.S. Schmidt, Oxford University Press, New York.
- Koop G., Osiewalski J., Steel M.F.J. 1994. *Hospital efficiency analysis through individual effects: A Bayesian approach*", CentER Discussion Paper 9447, Tilburg.
- Koop G., Osiewalski J., Steel M.F.J. 1997. *Bayesian efficiency analysis through individual effects: Hospital cost frontiers*, Journal of Econometrics, 79.
- Koop G., Steel M.F.J., Osiewalski J. 1995. *Posterior analysis of stochastic frontier models using Gibbs sampling*, Computational Statistics, 10.

- Lovell K.C. A. 1993. *Production frontiers and productive efficiency*, [w:] *The Measurement of Productive Efficiency — Techniques And Applications*, red. H.O. Fried, K.C.A. Lovell, S.S. Schmidt, Oxford University Press, New York.
- Marzec J., Osiewalski J. 2003. *Bayesowskie graniczne modele kosztów dla oddziałów banku. Wnioskowanie o efektywności kosztowej i jej determinantach*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, 628.
- Marzec J., Osiewalski J. 2008. *Bayesian inference on technology and efficiency of bank branches*, Bank i Kredyt 39/9.
- Meeusen W., van den Broeck J. 1977. *Efficiency estimation from Cobb-Douglas production functions with composed error*, International Economic Review, 8.
- O'Hagan A. 1994. *Bayesian Inference*, Edward Arnold, London.
- Osiewalski J. 1992. *Uogólnione niescentrowane współczynniki zwiększenia wariancji*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, 374.
- Osiewalski J. 2001. *Ekonometria bayesowska w zastosowaniach*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Osiewalski J., Osiewalska A. 2003. *Ocena efektywności kosztowej bibliotek akademickich na podstawie danych przekrojowo-czasowych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, 628.
- Osiewalski J., Osiewalska A. 2006. *Stochastyczna graniczna funkcja kosztu dla polskich bibliotek publicznych*, [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, red. A. Zeliaś, Akademia Ekonomiczna w Krakowie.
- Osiewalski J., Steel M.F.J. 1998. *Numerical tools for the Bayesian analysis of stochastic frontier models*, Journal of Productivity Analysis, 10.
- Osiewalski J., Wróbel-Rotter R. 1999. *Estymacja granicznych funkcji produkcji i wskaźników technicznej efektywności na podstawie danych przekrojowych*, Przegląd Statystyczny, 46.
- Osiewalski J., Wróbel-Rotter R. 2008. *Model ekonometryczny-narzędzie oceny efektywności spółek dystrybucyjnych kształtowanych w wyniku konsolidacji poziomej (skrót)*, Biuletyn Urzędu Regulacji Energetyki, 2 (58).
- Schmidt P., Sickles R.C. 1984. *Production frontiers and panel data*, Journal of Business and Economic Statistics, 2.
- Tierney L. 1994. *Markov chains for exploring posterior distributions (with discussion)*, Annals of Statistics, 22.
- Varian H.R. 1992. *Microeconomic Analysis*, W.W. Norton, New York.
- Wróbel-Rotter R. 2004. *Bayesowska analiza kosztu na podstawie stochastycznego granicznego modelu Leontiewa*, Przegląd Statystyczny, 51.
- Wróbel-Rotter R., Osiewalski J. 2002. *Bayesowski model efektów losowych w analizie efektywności kosztowej (na przykładzie elektrowni i elektrociepłowni polskich)*, Przegląd Statystyczny, 49.
- Zellner A. 1971. *An Introduction to Bayesian Inference in Econometrics*, J. Wiley, New York.

MONITOROWANIE PROCESÓW BINARNYCH ZA POMOCĄ KART KONTROLNYCH SUM SKUMULOWANYCH

ANDRZEJ IWASIEWICZ

Katedra Metod Statystycznych
Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego
30 -705 Kraków, ul. Gustawa Herlinga-Grudzińskiego 1
e-mail: andrzej.iwasiewicz@interia.pl

Praca była przedstawiona przez autora na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie, 16 grudnia 2008 r.

ABSTRACT

Andrzej Iwasiewicz, *Binary process monitoring by the cumulative sum control chart*. *Folia Aeconomica Cracoviensis* 2008–2009, 49–50: 71–90.

In the paper there have been presented some problems of the binary processes monitoring. There have been presented the monitoring of these processes by Shewhart control charts (control charts p and np) and cumulative sum control charts. Shewhart control charts are very useful if there are a possibility of a large shift in monitored parameters of process. A major disadvantage of a Shewhart control chart is that it uses only (first) the information about the process contained in the last observation and it rather ignores any information given by the entire sequence of empirical points. The cumulative control charts (cusum charts) are very effective alternatives to the Shewhart control charts. In the paper there have been presented the original modification of the cumulative sum control chart.

KEY WORDS — SŁOWA KLUCZOWE

monitoring, binary process, binary process monitoring, control chart, Shewhart control chart, cumulative sum control chart

monitorowanie, proces binarny, monitorowanie procesu binarnego, karta kontrolna, karta kontrolna Shewharta, karta kontrolna sum skumulowanych

1. PROCESY BINARNE

Pod pojęciem procesu binarnego należy rozumieć każdy proces dwustanowy, czyli taki proces, w przypadku którego badane zjawisko może znajdować się

tylko w dwóch, rozłącznych i wzajemnie wykluczających się stanach. Obserwowany proces może być albo z natury dwustanowy, albo obserwator tego procesu może dokonać dychotomizacji obrazu badanego zjawiska, sprowadzając go tym samym do postaci binarnej. Niezależnie od tego, z którym z tych przypadków mamy do czynienia, do opisu stanu obserwowanego zjawiska wykorzystuje się najczęściej zero-jedynkowe zmienne diagnostyczne. Wartości tych zmiennych (X) generowane są według następującej reguły:

$$\left. \begin{array}{l} X = 1 \text{ gdy badane zjawisko znajduje się w stanie wyróżnionym} \\ X = 0 \text{ gdy badane zjawisko nie znajduje się w stanie wyróżnionym.} \end{array} \right] \quad (1)$$

Monitorowanie procesu, nie tylko binarnego, podejmowane jest najczęściej w celu pozyskania informacji o średnim poziomie obserwowanego zjawiska, a także o jego zmienności. W przypadku procesów binarnych obydwa te zadania realizuje się poprzez śledzenie parametru:

$$p = P(X = 1), \quad (2)$$

charakteryzującego częstość pojawiania się tego stanu procesu, który w planie eksperymentu został potraktowany jako wyróżniony ($X = 1$). Przypomnijmy wszak, że w przypadku zero-jedynkowych zmiennych losowych wartość oczekiwana i wariancja dane są następującymi wzorami: $E(X) = p$, $D^2(X) = p \times (1 - p)$. Tak więc, znając wartość parametru p , albo uzyskane empirycznie punktowe lub przedziałowe oszacowania tej wartości, posiadamy informacje zarówno o średnim poziomie procesu, jak i o jego zmienności. Zauważmy też, że wariancja tej zmiennej losowej osiąga wartość maksymalną, a monitorowany proces najmniejszy stopień zdeterminowania przez kontrolowane uwarunkowania, gdy $p = 0,5$.

Przy przyjętych założeniach, formalnym modelem obserwowanego procesu $\{X_t\}$ jest ciąg zer ($X = 0$) i jedynek ($X = 1$), a częstość pojawiania się tych wartości określona jest przez parametr p . W konsekwencji, również uzyskiwane empirycznie liczbowe obrazy procesu są ciągami złożonymi z zer i jedynek¹. Są to zawsze ciągi skończone, podczas gdy monitorowany proces $\{X_t\}$ może być traktowany — w zależności od szczegółowych uwarunkowań — jako ciąg skończony lub nieskończony. Jeśli w kolejnych krokach postępowania kontrolnego mamy do czynienia z pojedynczymi ocenami natężenia obserwowanego zjawiska, to rezultatem przeprowadzonych badań jest ciąg:

$$x_1, x_2, x_3, \dots, x_t, \dots, x_k, \quad (3)$$

w którym x_t oznacza zrealizowaną wartość (zero albo jedynkę) zmiennej diagnostycznej X , w punkcie, albo przedziale, o numerze t ($t = 1, 2, 3, \dots, k$). Jeśli

¹ Ciągi takie nazywane są niekiedy słowami binarnymi.

natomiast w kolejnych krokach postępowania kontrolnego wielokrotnie powtarzana jest ocena natężenia badanego zjawiska, to każdej wartości t odpowiada pewien zbiór zer i jedynek, będących realizacjami zmiennej diagnostycznej X :

$$x_t = \{x_{t,1}, x_{t,2}, x_{t,3}, \dots, x_{t,i}, \dots, x_{tn_t}\}. \quad (4)$$

Ogólnym liczbowym obrazem monitorowanego procesu jest w takiej sytuacji ciąg tych zbiorów:

$$x_1, x_2, x_3, \dots, x_i, \dots, x_k. \quad (5)$$

Zbiory x_t są najczęściej próbkami losowymi, a n_t oznacza licznosc próbek pobranej do badania w chwili (przedziale) o numerze t . Odpowiednio zorganizowana analiza tych próbek umożliwia empiryczne pozyskiwanie informacji o stanie śledzonego procesu, które pozwalają wyznaczać punktowe i przedziałowe oceny parametru p zdefiniowanego wzorem (2), a także weryfikować hipotezy dotyczące tego parametru.

2. KARTY KONTROLNE

Procedury statystyczne znane pod nazwą kart kontrolnych (ang. *control chart*) są nieskomplikowanymi i bardzo użytecznymi narzędziami monitorowania procesów. Służą one do rejestrowania, wstępnego przetwarzania i analizy źródłowych informacji o stanie obserwowanego procesu, a ich zadaniem jest wykrywanie systematycznych (nielosowych) zmian w przebiegu tego procesu. Najczęściej, aczkolwiek nie zawsze, karta kontrolna jest elementem ujemnego (regulacyjnego) sprzężenia zwrotnego. W takiej sytuacji wykryta zmiana w przebiegu procesu, będąca nielosowym odchyleniem od zadanej normy, skutkuje emisją sygnału o rozregulowaniu procesu, a sygnał ten uruchamia odpowiednie działania regulacyjne. Takie usytuowanie karty kontrolnej w systemie monitorowania procesu nie jest jej jedynym zastosowaniem. Karta kontrolna może również funkcjonować w układzie dodatniego (deregulacyjnego) sprzężenia zwrotnego, a może być także wykorzystana do biernego śledzenia procesu, bez możliwości wpływania na jego przebieg. Prezentowane w literaturze przedmiotu karty kontrolne można klasyfikować według kilku kryteriów. Rozróżnia się więc przede wszystkim karty jednowymiarowe i wielowymiarowe. Podstawowe znaczenie praktyczne mają karty jednowymiarowe, stosowane w przypadku jednokryterialnej oceny badanych zjawisk. Takim kartom kontrolnym poświęcone są w całości przedstawione poniżej rozważania. Nie zmienia to oczywiście faktu, że bardzo często zachodzi potrzeba ocen wielokryterialnych. W takich sytuacjach definiuje się często jednowymiarowe, agregatowe zmienne diagnostyczne, charakteryzujące w sposób syntetyczny obserwowane, wielowymiarowe zjawisko. Niekiedy wykorzystuje się do tego celu zero-jedynkowe

zmienne losowe, a w konsekwencji obserwowany proces sprowadza się do postaci binarnej. Pełniejszy obraz monitorowanego procesu uzyskuje się stosując wielowymiarowe karty kontrolne, wśród których podstawową rolę odgrywają dwuwymiarowe karty kontrolne (T^2) Hotellinga (Montgomery, 2005). Podstawą klasyfikacji kart kontrolnych są także teoretyczne uwarunkowania ich funkcjonowania. Kierując się tym kryterium rozróżnia się karty kontrolne zaproponowane przez W. A. Shewharta, których teoretyczną bazę stanowi obecnie klasyczna teoria weryfikacji hipotez statystycznych² oraz karty kontrolne sum skumulowanych wykorzystujące sposób rozumowania zaproponowany przez A. Walda dla potrzeb analizy sekwencyjnej³. Innym, bardzo ważnym kryterium klasyfikacji kart kontrolnych jest ich zdolność do akceptacji monitorowanego procesu. Historycznie wcześniejsze są karty kontrolne, zarówno Shewharta, jak i sum skumulowanych, które nie posiadają tej zdolności. Potrzeby praktyki wymusiły jednak modyfikacje tych kart. Te zmodyfikowane karty kontrolne pozwalają akceptować obserwowany proces, a nie tylko dyskwalifikować go lub orzekać, że nie ma podstaw do dyskwalifikacji. Tak zmodyfikowane karty kontrolne są szczególnie użyteczne w sytuacjach, w których nie można uchylić się od odpowiedzi na pytanie: czy obserwowany proces przebiega prawidłowo, czy też wymaga korekty? Z taką sytuacją mamy do czynienia — na przykład — w działaniach związanych z audytem. Podstawowa część przedstawionych poniżej rozważań poświęcona jest monitorowaniu procesów z wykorzystaniem procedur sekwencyjnych. Omówienie funkcjonowania kart kontrolnych Shewharta stanowi niezbędne wprowadzenie do tego podstawowego nurtu rozważań.

3. KARTY KONTROLNE SHEWHARTA

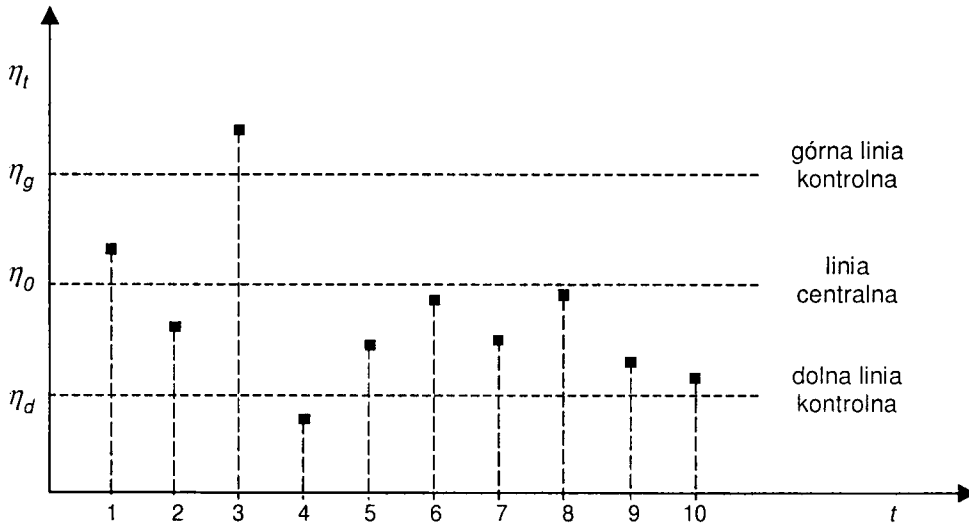
Najwcześniejsze historycznie, najbardziej znane i najczęściej stosowane są karty kontrolne pomysłu W.A. Shewharta. W latach dwudziestych ubiegłego stulecia przedstawił on i zrealizował koncepcję systemu wykorzystującego informacje generowane przez karty kontrolne do sterowania procesami wytwórczymi w celu zapewnienia wymaganego poziomu jakości wykonania⁴. Słowo

² Walter Andrew Shewhart (1891–1967), amerykański matematyk i statystyk, prekursor stosowania metod statystycznych dla potrzeb zarządzania jakością. Dodać należy, że w czasie gdy W. A. Shewhart pracował nad swoim systemem sterowania procesem z wykorzystaniem kart kontrolnych, teoria weryfikacji hipotez statystycznych nie była jeszcze ugruntowana. Powstawała ona niemal równoległe do prac W. A. Shewharta i była dziełem Jerzego Splawy-Neymana (1894–1981) oraz Egon Pearsona (1895–1980) (Rinne i Mittag, 1988).

³ Abraham Wald (1902–1950), statystyk amerykański (Rinne i Mittag, 1988)

⁴ Pierwszy taki system został uruchomiony 16 maja 1924 w Stanach Zjednoczonych (Juran, 1962).

„karta” (ang. *chart*) nieprzypadkowo znalazło się w nazwie omawianych procedur. Centralnym elementem karty kontrolnej jest odpowiednio skonstruowany diagram przeglądowy, służący do graficznego przedstawienia przebiegu śledzonego procesu. W przypadku kart kontrolnych zaproponowanych przez W. A. Shewharta, a także innych kart kontrolnych, diagram ten ma postać przedstawioną schematycznie na rycinie 1. Na osi poziomej odkładany jest numer próbki losowej pobranej z procesu, albo czas mierzony na skali porządkowej (t). Na osi pionowej odkłada się natomiast wartości obserwowanej charakterystyki z próby η_t , przy czym szczegółowa postać tej charakterystyki zależy od różnego rodzaju uwarunkowań, do których w szczególności należy zaliczyć cel badania oraz rodzaj monitorowanego procesu. Równoległe do osi t , na poziomie η_0 , wykreśla się linię centralną (ang. *center line*), odpowiadającą najczęściej, ale nie zawsze, średniemu poziomowi obserwowanego procesu. Poniżej i powyżej linii centralnej wykreśla się linie kontrolne, które mogą być bądź to granicami regulacji procesu (ang. *upper control limit*, *lower control limit*), bądź też mogą służyć do wykrywania korzystnych zmian w procesie. Na tak przygotowany diagram przeglądowy nanosi się kolejne punkty empiryczne (t , η_t), tworzące ciąg nazywany śladem procesu.



Ryc. 1. Diagram przeglądowy karty kontrolnej Shewharta

W zamysle W. A. Shewharta analiza tego ciągu miała być podstawą diagnoz dotyczących stanu obserwowanego procesu, a także decyzji sterujących jego przebiegiem. Operacje numeryczne miały być sprowadzone do niezbęd-

nego minimum. Przyczyna takiego potraktowania problemu wydaje się oczywista, zwłaszcza jeśli karty kontrolne miały funkcjonować w warunkach warsztatowych, na niskich szczeblach zarządzania. Należy pamiętać, że w latach dwudziestych ubiegłego stulecia, a także w kilku następnych dekadach, nie były dostępne te środki obliczeniowe, którymi dysponujemy obecnie. Był to świat bez kalkulatorów elektronicznych i bez komputerów. Dlatego też zapewne W. A. Shewhart zaproponował taki tok przetwarzania i analizy informacji źródłowych, w którym podstawowe oceny i decyzje formułowane są na podstawie graficznej analizy problemu. W świetle obecnej wiedzy o problemach wnioskowania statystycznego karty kontrolne Shewharta można rozważać jako sekwencje odpowiednich testów istotności. Dotyczy to jednak tylko takich sytuacji, gdy — z jakichkolwiek powodów — rola karty kontrolnej zredukowana jest do generowania i emisji tak zwanych punktowych sygnałów o rozregulowaniu monitorowanego procesu. Sygnał taki emitowany jest wówczas, gdy pojedynczy punkt empiryczny (t, η_t) znajdzie się poza obszarem ograniczonym przez linie kontrolne, będące granicami regulacji procesu. W sytuacji pokazanej na rycinie 1 sygnałami takimi są punkty $(3, \eta_3)$ oraz $(4, \eta_4)$. Każdy z tych sygnałów równoważny jest decyzji o odrzuceniu hipotezy zerowej za pomocą odpowiedniego testu istotności. Gdyby mianowicie na podstawie pojedynczego zbioru postaci (4) obliczyć wartość stosownej funkcji testowej i porównać ją z odpowiednią wartością krytyczną, to decyzja byłaby taka sama jak w przypadku sygnału punktowego wyemitowanego przez kartę kontrolną.

Analizując funkcjonowanie diagramu przeglądowego, przedstawionego na rycinie 1 nietrudno zauważyć, że karta kontrolna umożliwia nie tylko emisję sygnałów punktowych, ale stwarza także możliwość wykrywania objawów rozregulowania procesu poprzez analizę ciągu kolejnych punktów empirycznych (t, η_t) , tworzących ślad procesu. Celem tej analizy jest identyfikacja takich sekwencji (serii) punktów empirycznych, których pojawienie się w śladzie uregulowanego procesu, a więc przy założeniu prawdziwości hipotezy zerowej, jest bardzo mało prawdopodobne. Gdyby — na przykład — przyjąć, że w sytuacji przedstawionej na rycinie 1 obserwowana charakterystyka z próby (η) jest zmienną losową o normalnym rozkładzie prawdopodobieństwa, weryfikowana hipoteza zerowa ma postać $H_0 : E(\eta) = \eta_0$, a prawdopodobieństwo wyemitowania fałszywego sygnału o rozregulowaniu procesu ustalono na poziomie $\alpha = 0,05$, to ciąg sześciu punktów empirycznych $(5, \eta_5)$, $(6, \eta_6)$, $(7, \eta_7)$, $(8, \eta_8)$, $(9, \eta_9)$, $(10, \eta_{10})$ jest sygnałem sekwencyjnym (seryjnym), mimo że żaden z tych punktów, traktowany z osobna, nie wskazuje na rozregulowanie procesu. Wszystkie te punkty empiryczne leżą poniżej linii centralnej odpowiadającej wartości oczekiwanej zmiennej diagnostycznej. Jeśli są to niezależne oceny badanego zjawiska, to z symetrii rozkładu normalnego względem wartości średniej wynika, że prawdopodobieństwo przypadkowego ukształtowania się analizowanej sekwencji wynosi $(1/2)^6 = 0,0156 < \alpha = 0,05$. Jest to

konieczny i wystarczający warunek uznania analizowanego ciągu za sekwencyjny sygnał o rozregulowaniu procesu⁵.

4. MONITOROWANIE PROCESÓW BINARNYCH Z ZASTOSOWANIEM KART KONTROLNYCH SHEWHARTA

Skupmy uwagę na przypadku, gdy w kolejnych krokach postępowania kontrolnego pobierane są próbki losowe o licznosci n_t , a więc na przypadku, gdy empirycznym obrazem monitorowanego procesu jest ciąg postaci (5). Obliczając wartości:

$$z_t = \sum_{i=1}^{n_t} x_{t,i}, \quad (6)$$

uzyskujemy informację o liczbie elementów wyróżnionych ($X = 1$) w kolejnych ($t = 1, 2, 3, \dots, k$) zbiorach postaci (4). Jeśli próbki losowe pobierane do badania w kolejnych krokach postępowania kontrolnego są jednakowo liczne, to wartości z_t obliczone według wzoru (6) mogą być bezpośrednio wykorzystane do monitorowania procesu za pomocą karty kontrolnej. W takiej sytuacji należy posłużyć się kartą kontrolną z (ang. *control chart np*) Jeśli natomiast kolejno pobierane próbki mają różne licznosci, to wartości z_t obliczanych według wzoru (6) nie wykorzystuje się bezpośrednio do monitorowania procesu za pomocą klasycznej karty kontrolnej Shewharta. W celu zapewnienia porównywalności uzyskiwanych ocen cząstkowych, oblicza się wówczas wartości frakcji elementów wyróżnionych ($X = 1$), w kolejnych zbiorach postaci (4):

$$w_t = \frac{z_t}{n_t} \quad (7)$$

Mając te wartości można podjąć monitorowanie procesu za pomocą karty kontrolnej w (ang. *control chart p*).

W celu uściślenia dalszych rozważań ustalmy uwagę na przypadku, gdy stan wyróżniony ($X = 1$) jest niepożądanym (albo niekorzystnym) stanem monitorowanego procesu. W takiej sytuacji proces jest uregulowany (przebiega zgodnie z oczekiwaniami), gdy $p \leq p_0$, natomiast jest rozregulowany (nie przebiega on zgodnie z oczekiwaniami), gdy $p > p_0$, przy czym p_0 oznacza najgorszy, ale jeszcze dopuszczalny poziom procesu (ang. *Acceptable Process Level* — APL). Jeśli taki proces byłby monitorowany za pomocą wspomnianej powyżej karty kontrolnej z , albo karty kontrolnej w , to w każdym kroku postępowania kontrolnego weryfikowana byłaby hipoteza zerowa:

⁵ W normie PN-ISO 8258+AC1 wyróżniono siedem typów takich sygnałów seryjnych (sekwencyjnych), oprócz podstawowego sygnału punktowego. Problem ten omówiono również w pracy (Iwasiewicz i Stefanów, 2001).

$$H_0: p \leq p_0. \quad (8)$$

wobec hipotezy alternatywnej:

$$H_1: p > p_0. \quad (9)$$

W rozważanym przypadku granicami regulacji procesu byłyby tylko górne linie kontrolne, natomiast dolne linie kontrolne można by wykorzystać do wykrywania objawów korzystnych zmian w obserwowanym procesie⁶. W przypadku karty kontrolnej z granicą regulacji dana jest wzorem:

$$z_g = n \times p_0 + u_\alpha \times \sqrt{n \times p_0 \times (1 - p_0)}. \quad (10)$$

Stosując natomiast kartę kontrolną w , granicę regulacji procesu wyznaczamy według wzoru:

$$w_g = p_0 + u_\alpha \times \sqrt{\frac{p_0 \times (1 - p_0)}{n}}. \quad (11)$$

Nietrudno zauważyć, że obydwa te równania wynikają z prostego przekształcenia funkcji testowej odpowiedniego testu istotności, który można zastosować do weryfikacji sformułowanej powyżej hipotezy zerowej, w kolejnych krokach postępowania kontrolnego, traktowanych z osobna⁷. W konkretnym przypadku jest to test u , o następującej funkcji testowej:

$$u_0 = \frac{\frac{z}{n} - p_0}{\sqrt{\frac{p_0 \times (1 - p_0)}{n}}} = \frac{w - p_0}{\sqrt{\frac{p_0 \times (1 - p_0)}{n}}}. \quad (12)$$

W rozważanym powyżej przypadku emisja sygnału następuje wówczas, gdy $z_i > z_g$, albo $w_i > w_g$. Obydwa te sygnały równoważne są decyzji o odrzuceniu hipotezy zerowej (8) na korzyść hipotezy alternatywnej (9), gdy wartość u_0 obliczona na podstawie wzoru (12) spełnia nierówność $u_0 > u_\alpha$. W każdym z tych przypadków prawdopodobieństwo popełnienia błędu jest takie samo i wynosi α .

⁶ Wyczerpujący opis funkcjonowania kart kontrolnych Shewharta stosowanych w przypadku procesów binarnych przedstawiono w wielu podręcznikach; zob. np. Montgomery, 2005; Iwasiewicz, 1985, 1999, 2005.

⁷ Przegląd podstawowych testów istotności zainteresowany Czytelnik znajdzie w każdym obszerniejszym podręczniku statystyki; zob. np. Iwasiewicz i Paszek, 2004.

5. SEKWENCYJNE PROCEDURY MONITOROWANIA PROCESÓW

Możliwość identyfikacji i analizy nielosowych sekwencji punktów empirycznych (t, η_t) w śladzie procesu jest tą właściwością operacyjną karty kontrolnej Shewharta, która odróżnia ją od zwykłych sekwencji testów istotności. Możliwość ta zaowocowała pojawieniem się różnego rodzaju sekwencyjnych procedur kontrolnych. Należy tu przede wszystkim wymienić kartę kontrolną średnich ruchomych (ang. *moving average control chart* — MA chart). Karta ta może być stosowana albo jako samodzielna i odrębna procedura monitorowania procesu, albo jako dodatkowy segment standardowej karty kontrolnej Shewharta. W takiej sytuacji diagram przeglądowy standardowej karty kontrolnej wyposaża się dodatkowo w tor kontrolny służący do rejestrowania i analizy średnich ruchomych. Sygnały pojawiające się na tym torze nazywane są zwykle sygnałami z połączonych próbek. Takie postępowanie pozwala nie tylko na sprawniejsze wykrywanie sygnałów sekwencyjnych w śladzie procesu, ale wzmacnia również selektywność procedury, albowiem formułowane oceny oparte są na próbkach o zwielokrotnionej liczności. Na uwagę zasługuje rola, jaką karta kontrolna średnich ruchomych odgrywa w przypadku monitorowania procesów giełdowych. Ważnym instrumentem wykorzystywanym do monitorowania tych procesów są tak zwane wstęgi Bollingera (ang. *Bollinger bands*). Wstęga taka jest zmodyfikowaną kartą kontrolną średnich ruchomych, przy czym modyfikacje dotyczą sposobu wyznaczania zarówno linii centralnej, jak i linii kontrolnych. Wszystkie elementy procedury są tu wyznaczane na podstawie bieżąco pozyskiwanych informacji o stanie monitorowanego procesu, a nie na podstawie wartości postulowanych, jak w przypadku klasycznych kart kontrolnych Shewharta. Jak już wspomniano powyżej, teoretyczną podbudowę kart kontrolnych Shewharta stanowi teoria testów istotności. Testy te — w standardowej postaci — nie służą do przyjmowania weryfikowanych hipotez zerowych. Za pomocą testu istotności można albo odrzucić hipotezę zerową przy określonym prawdopodobieństwie popełnienia pomyłki (α), albo orzec, że nie ma podstaw do jej odrzucenia. Nie można natomiast przyjąć (zaakceptować) hipotezy zerowej, albowiem prawdopodobieństwo popełnienia błędu drugiego rodzaju (β), polegającego na przyjęciu fałszywej hipotezy zerowej, pozostaje poza bezpośrednią kontrolą. Każdy standardowy test istotności można wprawdzie przekształcić w taki sposób, by istniała możliwość przyjmowania weryfikowanych hipotez zerowych, ale znacznie efektywniejszą metodą konstruowania takich testów jest analiza sekwencyjna. Jest to metoda efektywniejsza przede wszystkim w sensie niezbędnej liczności próbki, a więc również w sensie kosztów monitorowania procesu. Teoria sekwencyjnych procedur weryfikacji hipotez, zaproponowana przez A. Walda, stała się obecnie teoretyczną podstawą dla nowego rodzaju kart kontrolnych, a mianowicie kart kontrolnych sum skumulowanych (ang. *cumulative-sum control chart, cusum control chart*), mimo że początkowe prace w tym zakresie nie odwoływały się do tej teorii.

6. PROCEDURY KONTROLNE UMOŻLIWIAJĄCE AKCEPTACJĘ MONITOROWANEGO PROCESU BINARNEGO

W celu utrzymania ciągłości rozumowania pozostaniemy przy przyjętych powyżej założeniach w odniesieniu merytorycznej natury obserwowanego procesu binarnego. W dalszym ciągu będziemy mianowicie zakładać, że wyróżniony stan badanego zjawiska ($X = 1$) jest stanem nacechowanym negatywnie i że — w konsekwencji — zbiór dopuszczalnych poziomów procesu (p) ma postać przedziału $[0; p_0]$. Tak więc, proces jest uregulowany (przebiega zgodnie z oczekiwaniami) gdy $p \leq p_0$, natomiast wymaga on regulacji (korekty) gdy $p > p_0$. Jeśli — w rozważanej sytuacji — do monitorowania procesu wykorzystywana jest procedura bez możliwości akceptacji procesu, a więc np. któraś ze wspomnianych powyżej klasycznych kart kontrolnych Shewharta (z albo w), to weryfikowana hipoteza zerowa ma postać (8), natomiast treść hipotezy alternatywnej wyrażona jest nierównością (9). Jeśli chcemy uzyskać możliwość akceptacji monitorowanego procesu, to hipotezy te muszą być odpowiednio zmodyfikowane. Sformułujemy je następująco:

$$H_0: p = p_0 \quad (13)$$

$$H_1: p = p_0 + \Delta p = p_1, \quad (14)$$

przy czym $\Delta p > 0$.

Każda z tych hipotez może być odrzucona na korzyść hipotezy konkurencyjnej. Odrzucenie hipotezy zerowej (13) na korzyść hipotezy alternatywnej (14) skutkuje emisją sygnału o rozregulowaniu procesu, a prawdopodobieństwo emisji sygnału fałszywego nie przekracza przyjętej wstępnie, dowolnie małej, dodatniej wartości α . Odrzucenie hipotezy alternatywnej (14) na korzyść hipotezy zerowej (13) oznacza akceptację monitorowanego procesu. Prawdopodobieństwo fałszywego sygnału nie przewyższa w tym przypadku wstępnie przyjętej, dowolnie małej, dodatniej wartości β .

Zauważmy, że — analogicznie jak poprzednio — monitorowany proces traktujemy jako rozregulowany, gdy $p > p_0$. Godzimy się jednak na to, by wystarczająco pewny sygnał o rozregulowaniu pojawiał się wówczas, gdy:

$$p > p_0 + \Delta p = p_1. \quad (15)$$

Przedział $(p_0; p_1)$, o długości $p_1 - p_0 = \Delta p > 0$, należy interpretować jako przedział decyzji niepewnych. Są one niepewne w tym sensie, że w przedziale $(p_0; p_1)$ mogą pojawić się sygnały o akceptacji lub rozregulowaniu monitorowanego procesu z prawdopodobieństwami znacznie różniącymi się od przyjętych wartości α i β . Wartość różnicy $p_1 - p_0 = \Delta p$ deklarowana jest na podstawie przesłanek merytorycznych, a nie statystycznych. Jest to takie odchylenie rzeczywistego poziomu procesu (p) od zadanej wartości p_0 , które merytorycznie,

a nie statystycznie, można uznać za nieistotne. Konieczność definiowania przedziału decyzji niepewnych jest naturalną konsekwencją ciągłości funkcji mocy albo funkcji operacyjno-charakterystycznej omawianej procedury, w całym przedziale zmienności parametru p , czyli w przedziale $[0; 1]$ ⁸. Ciągłość tych funkcji sprawia, że w badaniach niewyczerpujących jakakolwiek skończona liczność próbek nie wystarcza do odróżnienia stanu $p \leq p_0$ od stanu $p > p_0$, przy wystarczająco małych wartościach α i β . Również w badaniach wyczerpujących odróżnienie tych stanów jest wątpliwe ze względu na możliwość występowania błędów kwalifikacji (diagnozowania), zob. Iwasiewicz, 2005–2006.

Sformułowane powyżej uwagi o konstruowaniu procedur kontrolnych umożliwiających akceptację monitorowanego procesu dotyczą zarówno procedur uzyskiwanych w rezultacie odpowiedniej modyfikacji kart kontrolnych typu Shewharta, jak i procedur sekwencyjnych⁹. Przedstawione poniżej rozważania dotyczą w całości procedur sekwencyjnych, w przypadku których istnieje możliwość istotnego ograniczenia liczności próbki, co jest bardzo ważne z punktu widzenia kosztów monitorowania procesu. W przypadku zmodyfikowanych kart kontrolnych Shewharta możliwości takie są bardzo ograniczone.

7. KLASYCZNA PROCEDURA SEKWENCYJNA

Przedstawione poniżej rozważania dotyczą sytuacji, gdy w kolejnych krokach postępowania kontrolnego badaniu poddawane są pojedyncze elementy śledzonego procesu. Rezultatem badania jest wówczas ciąg zer i jedynek postaci (3). Ciąg ten jest stopniowo wydłużany, stosownie do potrzeb badania, a jego każdorazowa długość (k) jest tożsama z licznością próbki ($k = n$).

Jeśli ustalone są wartości α , β , p_0 i p_1 , to warunek kontynuowania obserwacji, czyli wydłużania ciągu (3), przedstawia się następująco:

$$\frac{\beta}{1 - \alpha} < \frac{\prod_{t=1}^n p_1^{x_t} q_1^{1-x_t}}{\prod_{t=1}^n p_0^{x_t} q_0^{1-x_t}} < \frac{1 - \beta}{\alpha}, \quad (16)$$

przy czym $q_0 = 1 - p_0$ oraz $q_1 = 1 - p_1$, natomiast x_t jest zrealizowaną wartością (zerem albo jedynką) zmiennej diagnostycznej X , w kolejnym, t -tym kroku postępowania kontrolnego.

Jeśli obserwowaną charakterystyką z próby jest suma zrealizowanych wartości zmiennej diagnostycznej X , to w kolejnych krokach postępowania oblicza się wartości:

⁸ Wyczerpujące omówienie tego problemu zainteresowany Czytelnik znajdzie w pracy: Iwasiewicz, 1999.

⁹ Opis takich modyfikacji przedstawiono w pracach: Major, 1997; Iwasiewicz, 2001.

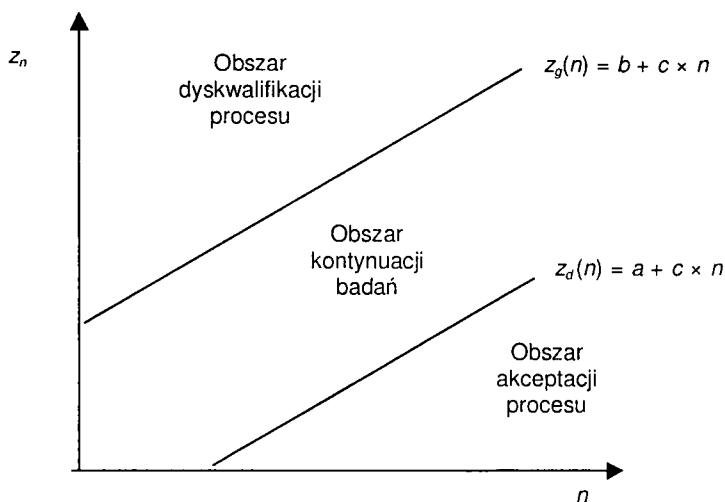
$$z_t = \sum_{i=1}^n x_i, \quad (17)$$

przy czym $n = 1, 2, 3, \dots$, natomiast $t = 1, 2, 3, \dots, n$.

W graficznej wersji omawianej procedury obliczanie tej sumy przebiega według następującego schematu:

$$\begin{aligned} t = 1: & \quad z_1 = x_1 \\ t = 2: & \quad z_2 = z_1 + x_2 \\ t = 3: & \quad z_3 = z_2 + x_3 \\ & \quad \dots \\ t = n: & \quad z_n = z_{n-1} + x_n. \end{aligned} \quad (18)$$

Obliczone w ten sposób wartości sumy (17) nanosi się na diagram przeglądkowy przedstawiony schematycznie na rycinie 2.



Ryc. 2. Diagram przeglądkowy klasycznej sekwencji procedury kontrolnej

Równania linii kontrolnych zaznaczonych na tym rysunku wynikają z nierówności (16). Odpowiednie przekształcenia przedstawiają się następująco:

$$\frac{\beta}{1-\alpha} < \frac{p_1^{z_n} q_1^{n-z_n}}{p_0^{z_n} q_0^{n-z_n}} < \frac{1-\beta}{\alpha}, \quad (19)$$

$$\frac{\beta}{1-\alpha} < \left(\frac{p_1}{p_0}\right)^{z_n} \left(\frac{q_1}{q_0}\right)^{n-z_n} < \frac{1-\beta}{\alpha}. \quad (20)$$

Po zlogarytmowaniu tej nierówności i dalszych przekształceniach warunków kontynuowania badań można zapisać następująco:

$$z_d(n) = a + c_n < z_n < b + c_n = z_g(n), \quad (21)$$

gdzie:

$$c = \frac{\ln \frac{q_0}{q_1}}{\ln \frac{p_1 q_0}{p_0 q_1}} > 0, \quad (22)$$

$$a = \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} < 0, \quad (23)$$

$$b = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} > 0. \quad (24)$$

Jeśli spełniona jest nierówność (21), to przechodzi się do badania następnego elementu monitorowanego procesu, bez podejmowania jakichkolwiek decyzji w odniesieniu do hipotez (13) i (14). Jeśli:

$$z_n \leq z_d(n) = a + c \times n, \quad (25)$$

to przyjmujemy hipotezę zerową (13), a prawdopodobieństwo zdarzenia losowego polegającego na tym, że w rzeczywistości prawdziwa jest hipoteza alternatywna (14) nie przekracza przyjętej wartości β . Jeśli natomiast:

$$z_n \geq z_g(n) = b + c \times n, \quad (26)$$

to przyjmujemy hipotezę alternatywną (14), a prawdopodobieństwo zdarzenia losowego polegającego na tym, że w rzeczywistości prawdziwa jest hipoteza zerowa (13) nie przekracza przyjętej wartości α . Przedstawione postępowanie można realizować w formie algorytmu graficznego, za pomocą diagramu przeglądowego przedstawionego na rycinie 2, albo algorytmu numerycznego, przy wykorzystaniu odpowiedniej tablicy decyzyjnej.

8. KARTA KONTROLNA SUM SKUMULOWANYCH

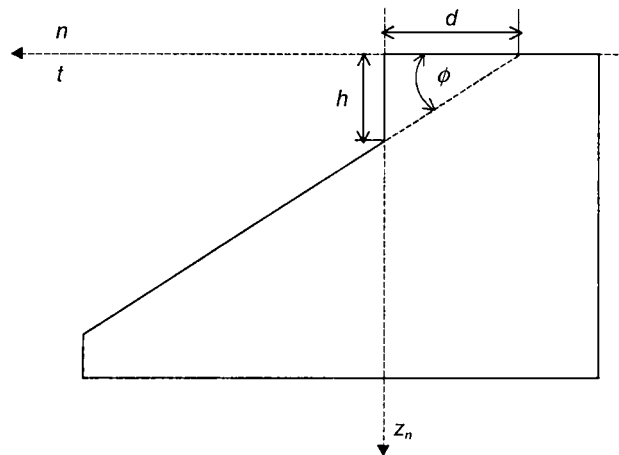
Kartę kontrolną sum skumulowanych można rozważać jako klasyczną procedurę sekwencyjną realizowaną wstecznie. W graficznej wersji procedury, na diagram przeglądowy wykreślony w układzie współrzędnych $(n0z_n)$ nanosi się kolejne sumy skumulowane z_n obliczane według następującego schematu:

$$\begin{aligned}
 t = 0: & \quad z_0 = x_0 = 0 \\
 t = 1: & \quad z_1 = z_0 + x_1 \\
 t = 2: & \quad z_2 = z_1 + x_2 \\
 t = 3: & \quad z_3 = z_2 + x_3 \\
 & \quad \dots \\
 t = n: & \quad z_n = z_{n-1} + x_n.
 \end{aligned}
 \tag{27}$$

W celu zachowania ciągłości rozumowania pozostajemy więc przy założeniu, że podstawą analizy jest ciąg punktów empirycznych postaci (3). Oznacza to, że w każdym kroku postępowania pobierana jest do badania jednoelementowa próbka losowa ($n_t = 1$). W konsekwencji, każdorazowo mamy $t = n$. Należy podkreślić, że do sekwencji (27) włączony jest początek układu współrzędnych, co nie było konieczne w przypadku klasycznej procedury sekwencyjnej; zob. wzór (18). Dołączenie wartości $x_0 = 0$ do sekwencji sum skumulowanych nie zmienia oczywiście wartości tej sumy:

$$z_n = \sum_{t=0}^n x_t = \sum_{t=1}^n x_t. \tag{28}$$

Każdy punkt (n, z_n) , kończący w kolejnym kroku postępowania kontrolnego sekwencję (27), traktuje się jako początek obróconego o 180° układu współrzędnych. W takiej sytuacji wykreślanie linii kontrolnych jest rozwiązaniem nieracjonalnym. Linie kontrolne zastępowane są maskownicą przesuwaną na płaszczyźnie rysunku, w miarę wydłużania się śladu procesu i sekwencji (27). W standardowych wersjach tej procedury obszar kontynuacji badań łączony jest z obszarem przyjęć hipotezy zerowej. W konsekwencji, do wyznaczania



Ryc. 3. Maskownica stosowana w procedurze bez możliwości akceptacji procesu

parametrów maskownicy, której schemat pokazano na rycinie 3, wykorzystuje się tylko górną linię kontrolną klasycznego schematu sekwencyjnego¹⁰.

Uwzględniając podane powyżej wzory (21), (22) i (24) równanie górnej linii kontrolnej w klasycznym schemacie sekwencyjnym można zapisać następująco:

$$z_s(n) = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} + \frac{\ln \frac{q_0}{q_1}}{\ln \frac{p_1 q_0}{p_0 q_1}} \times n. \quad (29)$$

W konsekwencji połączenia obszaru przyjęć hipotezy zerowej z obszarem kontynuacji badań przyjmuje się $\beta = 0$. Mamy więc:

$$z_s(n) = \frac{\ln \frac{1}{\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} + \frac{\ln \frac{q_0}{q_1}}{\ln \frac{p_1 q_0}{p_0 q_1}} \times n. \quad (30)$$

W celu uzyskania wartości parametru d maskownicy należy wyznaczyć pierwiastek tego równania:

$$n^* = \frac{\ln \alpha}{\ln \frac{q_0}{q_1}}.$$

Uwzględniając fakt, że parametry maskownicy wyznaczane są w układzie współrzędnych obróconym o 180° mamy:

$$d = -n^* > 0. \quad (31)$$

Drugi parametr maskownicy, czyli tangens kąta nachylenia czynnej krawędzi maskownicy względem osi odciętych ($c = \text{tg} \varphi$) dany jest wzorem (22).

Mając wartości parametrów c i d można skonstruować algorytm numeryczny, formalnie równoważny przedstawionemu powyżej algorytmowi graficznemu, a jednocześnie nie wymagający sporządzenia maskownicy i odpowiedniego diagramu przeglądowego. Zastosowanie algorytmu numerycznego wymaga natomiast skonstruowania odpowiedniej tablicy obliczeniowej i wyznaczenia wartości dodatkowego parametru h , pełniącego rolę wartości krytycznej dla sum skumulowanych, obliczanych w kolejnych krokach postępowania kontrolnego. Wartość parametru h oblicza się według wzoru:

¹⁰ Wyczerpujące omówienie funkcjonowania standardowych kart kontrolnych sum skumulowanych zainteresowany Czytelnik znajdzie w pracach: Montgomery, 2005; Iwasiewicz 1985, 1999.

$$h = c \times d = -\frac{\ln \alpha}{\ln \frac{q_0}{q_1}} \times \frac{\ln \frac{q_0}{q_1}}{\ln \frac{p_1 \times q_0}{p_0 \times q_1}} = -\frac{\ln \alpha}{\ln \frac{p_1 \times q_0}{p_0 \times q_1}}. \quad (32)$$

Dalsze postępowanie przebiega według przedstawionego poniżej schematu:

1. Każdą uzyskaną empirycznie wartość x_t (0 albo 1) porównuje się z wartością parametru c . Jeśli $x_t \leq c$, a więc jeśli $x_t = 0$, to nie podejmuje się żadnych działań, a po przejściu do następnego, losowo wybranego punktu na osi czasu generuje się kolejną wartość x_t ¹¹. Jeśli natomiast spełniona jest nierówność $x_t > c$, a więc jeśli $x_t = 1$, to uruchamia się licznik indeksu operacyjnego ($j = 1$) i rozpoczyna się obserwację sekwencji, która może doprowadzić do odrzucenia hipotezy zerowej (13) na korzyść hipotezy alternatywnej (14).

2. W tym celu oblicza się wartość sumy skumulowanej:

$$D_r = \sum_{j=1}^r (x_{t_j} - c), \quad (33)$$

gdzie: t — indeks bieżący, j — indeks operacyjny, r = największa w danym momencie wartość indeksu operacyjnego ($r = 1, 2, 3, \dots$).

3. Sygnał o rozregulowaniu monitorowanego procesu emitowany jest wówczas, gdy spełniona jest nierówność:

$$D_r \geq h. \quad (34)$$

Oznacza to odrzucenie hipotezy zerowej (13) na korzyść hipotezy alternatywnej (14), a prawdopodobieństwo zdarzenia losowego polegającego na tym, że jest to sygnał fałszywy nie przekracza przyjętej wartości α . W konsekwencji powinny być podjęte działania zmierzające do przywrócenia procesowi poprawnego przebiegu. Emisja sygnału o rozregulowaniu procesu kończy obserwowaną sekwencję. Następuje zerowanie licznika indeksu operacyjnego ($j = 0$).

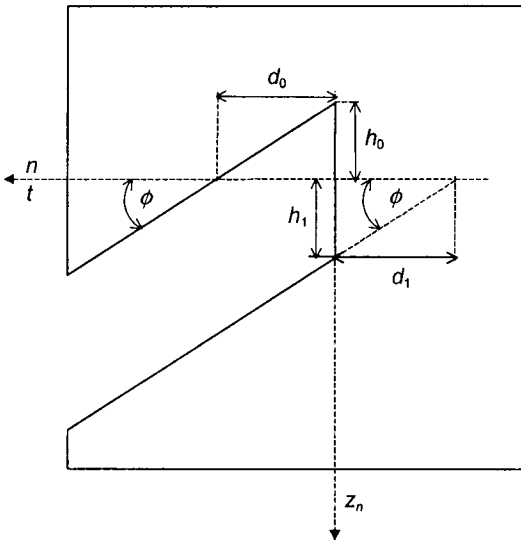
4. Śledzenie rozpoczętej sekwencji kończy się również wówczas gdy:

$$D_r \leq 0. \quad (35)$$

W takiej sytuacji również następuje zerowanie licznika indeksu operacyjnego ($j = 0$), jednak bez podejmowania działań korygujących przebieg procesu.

Przedstawione powyżej algorytmy, zarówno graficzny, jak i numeryczny, można zmodyfikować w taki sposób, by omawiana procedura umożliwiała akceptację monitorowanego procesu. W tym celu należy zrezygnować z zastosowanego powyżej łączenia obszaru akceptacji procesu z obszarem kontynuowa-

¹¹ Wygodnym, a niezbyt często stosowanym sposobem losowego wyboru punktów na osi czasu jest zastosowanie liczb złotych (Iwasiewicz i Paszek, 2004).



Ryc. 4. Maskownica stosowana w procedurze z możliwością akceptacji procesu

nia obserwacji. W konsekwencji, maskownica stosowana w graficznej wersji procedury ma kształt pokazany schematycznie na rycinie 4. Szczegółowy kształt tej maskownicy określony jest przez trzy — zaznaczone na rysunku — parametry, a mianowicie: d_0 , d_1 oraz $c = \operatorname{tg} \phi$. Dla potrzeb algorytmu numerycznego wyznacza się dodatkowo h_0 oraz h_1 . Jeśli wartości p_0 i p_1 są ustalone, to parametr c ma tu taką samą wartość jak w standardowej procedurze, bez możliwości akceptacji procesu, albowiem nie zależy ona od przyjętych wartości α i β . Parametry d_0 i h_0 dotyczą tego segmentu procedury, który umożliwia przyjęcie hipotezy zerowej H_0 , a więc akceptację śledzonego procesu w tym sensie, że $p \leq p_0$. Prawdopodobieństwo zdarzenia losowego polegającego na tym, że w rzeczywisty poziom procesu p spełnia nierówność $p > p_1$, nie przekracza przyjętej wstępnie, dowolnie małej, dodatniej wartości β . Parametry d_1 i h_1 dotyczą natomiast tego segmentu procedury, który umożliwia przyjęcie hipotezy alternatywnej H_1 , a więc dyskwalifikację monitorowanego procesu. Przyjmując H_1 orzekamy, że $p > p_1$, a prawdopodobieństwo zdarzenia losowego polegającego na tym, że w rzeczywistości proces jest uregulowany ($p \leq p_0$), nie przekracza przyjętej wstępnie, dowolnie małej, dodatniej wartości α . Parametr d_0 uzyskujemy w rezultacie przedstawionych poniżej przekształceń równania dolnej linii kontrolnej w klasycznej procedurze sekwencyjnej (zob. ryc. 2). Równanie to przedstawia się następująco:

$$z_d(n) = \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} + \frac{\ln \frac{q_0}{q_1}}{\ln \frac{p_1 q_0}{p_0 q_1}} \times n. \quad (36)$$

Chcąc uzyskać równanie parametru d_0 należy wyznaczyć pierwiastek (miejsce zerowe) powyższego równania (n_0). Po niezbędnych przekształceniach mamy:

$$\ln \frac{\beta}{1-\alpha} + \ln \frac{q_0}{q_1} \times n = 0, \quad (37)$$

$$n_0 = \frac{-\ln \frac{\beta}{1-\alpha}}{\ln \frac{q_0}{q_1}} > 0. \quad (38)$$

W konsekwencji:

$$d_0 = \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{q_0}{q_1}} < 0. \quad (39)$$

Równanie parametru h_0 — wykorzystywanego w numerycznej wersji procedury — przedstawia się następująco:

$$h_0 = c \times d_0 = \frac{\ln \frac{q_0}{p_1 q_1}}{\ln \frac{p_1 q_0}{p_0 q_1}} \times \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{q_0}{q_1}} = \frac{\ln \frac{\beta}{1-\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} < 0. \quad (40)$$

W analogiczny sposób uzyskuje się równania parametrów d_1 i h_1 . Mamy mianowicie:

$$d_1 = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{q_0}{q_1}} > 0, \quad (41)$$

$$h_1 = c \times d_1 = \frac{\ln \frac{q_0}{p_1 q_1}}{\ln \frac{p_1 q_0}{p_0 q_1}} \times \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{q_0}{q_1}} = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_1 q_0}{p_0 q_1}} > 0. \quad (42)$$

Dysponując wartościami parametrów c , h_0 i h_1 można skonstruować numeryczny algorytm weryfikacji hipotez (13) i (14), w odniesieniu do monitorowanego procesu. Analogicznie jak poprzednio, każdą uzyskaną empirycznie wartość x_t (0 albo 1) porównuje się z wartością parametru $c = \tan \phi$. Wynik tego porównania wyznacza drogę dalszego postępowania.

Analiza sekwencji, która może doprowadzić do przyjęcia hipotezy zerowej (13), a więc do stwierdzenia, że monitorowany proces jest uregulowany, przebiega według przedstawionego poniżej schematu.

1. Śledzenie tej sekwencji rozpoczyna się wówczas gdy pojawi się wartość $x_t < c$, a więc wartość $x_t = 0$. Uruchamia się wówczas licznik indeksu operacyjnego ($i = 1$).

2. W dalszym postępowaniu oblicza się wartość sumy skumulowanej:

$$D_{0,s} = \sum_{i=1}^s (x_{ti} - c), \quad (43)$$

gdzie: t — indeks bieżący, i — indeks operacyjny, s — największa w danym momencie wartość indeksu operacyjnego ($s = 1, 2, 3, \dots$).

3. Przyjęcie hipotezy zerowej (13) oraz zakończenie śledzenia danej sekwencji punktów empirycznych i zerowanie licznika indeksu operacyjnego ($i = 0$) następuje wówczas, gdy:

$$D_{0,s} \leq h_0, \quad (44)$$

przy czym $h_0 < 0$, zob. wzór (40). Przyjmując hipotezę zerową (13) i akceptując — w konsekwencji — przebieg śledzonego procesu, z prawdopodobieństwem β narażamy się na ryzyko, że w rzeczywistości prawdziwa jest hipoteza alternatywna (14).

4. Analizę obserwowanej sekwencji kończy się również wtedy, kiedy zostanie spełniona nierówność:

$$D_{0,s} \geq 0. \quad (45)$$

W takiej sytuacji również zeruje się licznik indeksu operacyjnego ($i = 0$).

Analiza sekwencji, która może doprowadzić do przyjęcia hipotezy alternatywnej (14), a więc do stwierdzenia, że monitorowany proces jest rozregulowany, przebiega według przedstawionego poniżej schematu.

1. Śledzenie tej sekwencji rozpoczyna się wówczas gdy pojawi się wartość $x_t > c$, a więc wartość $x_t = 1$. Uruchamia się wówczas licznik indeksu operacyjnego ($j = 1$).

2. W dalszym postępowaniu oblicza się wartości sumy skumulowanej według wzoru (33). Sumę tę oznaczymy obecnie symbolem $D_{1,r}$ dla odróżnienia od analogicznej sumy w algorytmie bez możliwości akceptacji monitorowanego procesu.

3. Przyjęcie hipotezy alternatywnej (14) i zakończenie śledzenia danej sekwencji punktów empirycznych następuje wówczas, gdy spełniona jest nierówność:

$$D_{1,r} \geq h_1. \quad (46)$$

W konsekwencji emitowany jest sygnał o rozregulowaniu monitorowanego procesu, a prawdopodobieństwo zdarzenia losowego polegającego na tym,

że jest to sygnał fałszywy nie przekracza przyjętej wartości α . W konsekwencji powinny być podjęte działania zmierzające do przywrócenia procesowi poprawnego przebiegu. Emisja sygnału o rozregulowaniu procesu kończy obserwowaną sekwencję. Następuje zerowanie licznika indeksu operacyjnego ($j = 0$).

4. Śledzenie rozpoczętej sekwencji kończy się również wówczas, gdy:

$$D_{1,r} \leq 0. \quad (47)$$

W takiej sytuacji również następuje zerowanie licznika indeksu operacyjnego ($j = 0$), jednak bez podejmowania działań korygujących przebieg procesu.

Należy podkreślić, że wyróżnione powyżej dwie konkurencyjne sekwencje punktów empirycznych nie są rozłączne. Nie są one rozłączne w tym sensie, że obydwie mogą przebiegać jednocześnie. I tak np. sekwencja mogąca doprowadzić do przyjęcia hipotezy zerowej H_0 — w konsekwencji — do akceptacji monitorowanego procesu, rozpoczyna się od wartości $x_i = 0$, ale nie jest przecież wykluczone, że w czasie jej trwania pojawi się punkt empiryczny o wartości $x_i = 1$. Punkt taki rozpoczyna nową sekwencję, mogącą doprowadzić do przyjęcia hipotezy alternatywnej H_1 i dyskwalifikacji procesu, ale jednocześnie nie przerywa on wcześniej rozpoczętej, konkurencyjnej sekwencji dopóty, dopóki nie zostanie spełniona nierówność (45). To samo dotyczy oczywiście sekwencji rozpoczynającej się od obserwacji empirycznej o wartości $x_i = 1$.

BIBLIOGRAFIA

- Achelis S.B. 1998. *Analiza techniczna od a do z*, tłum. z j. ang., Oficyna Wydawnicza LT & P, Warszawa.
- Iwasiewicz A. 1985. *Statystyczna kontrola jakości w toku produkcji; systemy i procedury*, PWN, Warszawa.
- Iwasiewicz A. 1999. *Zarządzanie jakością*, Wydawnictwo Naukowe PWN, Warszawa–Kraków.
- Iwasiewicz A. 2001. *Karty kontrolne Shewharta z możliwością akceptacji procesu*, Zeszyty Naukowe Uniwersytetu Szczecińskiego, 320, 35–48.
- Iwasiewicz A. 2005. *Zarządzanie jakością w przykładach i zadaniach*, Śląskie Wydawnictwa Naukowe Wyższej Szkoły Zarządzania i Nauk Społecznych w Tychach, Tychy.
- Iwasiewicz A. 2005–2006. *Monitorowanie procesów binarych*, Folia Oeconomica Cracoviensia, 46–47, 103–116.
- Iwasiewicz A., Paszek A. 2004. *Statystyka z elementami statystycznych metod monitorowania procesów*, wyd. IV poprawione, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Iwasiewicz A., Stefanów P. 2000. *Wykorzystanie testów wzorca przebiegu w statystycznym sterowaniu procesami*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, 874, 199–207.
- Juran J.M. 1962. *Handbook of quality*, McGraw Hill, New York.
- Major M. 1997. *Sterowanie procesem za pomocą kartkontrolnych sum skumulowanych*, Materiały konferencyjne z I Krajowej Konferencji Naukowej Materiałoznawstwo — Odlewnictwo, t. 3, Jakość, Kraków.
- Montgomery D.C. 2005. *Introduction to statistical quality control, fifth edition*, John Wiley & sons, inc.
- PN-ISO 8258+AC1 1996. *Karty kontrolne Shewharta*.
- Rinne H., Mittag H.J. 1988. *Statistische Methoden der Qualitätssicherung*, Fernuniversitaet — Gesamthochschule — in Hagen.

UWARUNKOWANIA POMIARU WSPÓLZALEŻNOŚCI CECH

TADEUSZ GRABIŃSKI

Katedra Finansów
Uniwersytet Ekonomiczny w Krakowie
PL 31-510 Kraków, ul. Rakowicka 27
email: tg@ae.krakow.pl

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych PAN w dniu 20 stycznia 2009 r. przez autora.

ABSTRACT

Tadeusz Grabiński, *Problems of measuring statistical dependence of variables*, Folia Oeconomica Cracoviensia, 2008–2009, 49–50: 91–107.

The paper presents propositions of modification in correlation and regression linear analysis. It is a result of problems in interpretation of statistical dependence measures in case where data are referring to aggregate units (e.g. countries). The paper shows results of research made at the beginning of 2000 for 81 countries by R. Lynn and T. Vanhanen who show that IQ is one of important factors contributing to differences in wealth of nations and rates of economic growth.

KEY WORDS — SŁOWA KLUCZOWE

measures of dependence, analysis of correlation and regression, wealth of nation, IQ
miary współzależności, analiza korelacji i regresji, bogactwo narodów, iloraz inteligencji

1. WPROWADZENIE

Podstawą badań statystycznych są informacje opisujące analizowane elementy zbiorowości (próby) z punktu widzenia wybranej cechy (jednej lub więcej). W każdej analizie należy zdefiniować jednostki badania oraz określić cechy charakteryzujące te jednostki.

Wyróżnia się jednostki elementarne oraz agregatowe. Trudno podać precyzyjne rozróżnienie tych pojęć, gdyż zależy ono od kontekstu

analizy. Przykładem jednostek elementarnych jest osoba, firma, produkt, gmina i w tym kontekście odpowiednimi jednostkami agregatowymi są: gospodarstwa domowe dla osób; branża dla firm; region, województwa, powiaty dla gmin itp.

To rozróżnienie nie zawsze jest brane pod uwagę w analizach ekonometrycznych i może prowadzić do kłopotów w interpretacji uzyskanych wyników. W wielu przypadkach mamy do czynienia z jednostkami agregatowymi, które traktujemy z jednakową wagą, podczas gdy nie zawsze jest to uzasadnione wielkością tych jednostek.

Dodatkowym elementem, na który należy zwrócić szczególną uwagę w badaniach są miana (jednostki pomiarowe) analizowanych cech. Mogą to być miana naturalne (zł, kg, osoby) lub złożone o charakterze wskaźników (dochód narodowy na osobę, zarobek tygodniowy na 1 zatrudnionego).

Przedmiotem niniejszego artykułu jest przedstawienie problemów związanych z istotą jednostek analizy i ich mian w przypadku pomiaru współzależności cech. Rozważania ograniczone zostaną do najprostszych metod oceny współzależności — liniowych współczynników korelacji i regresji jednej zmiennej, ale można je rozszerzyć na bardziej zaawansowane miary.

Dla ułatwienia przyjmuje się najprostszy przypadek analizy, w którym mamy do czynienia z danymi statycznymi. Punktem wyjścia rozważań są więc informacje liczbowe (skala ilorazowa) dotyczące n -elementowego zbioru jednostek, opisujące je z punktu widzenia dwóch cech X i Y w wybranym momencie czasowym:

$$\begin{aligned} \{X\} &= [x_1, x_2, \dots, x_n] \\ \{Y\} &= [y_1, y_2, \dots, y_n]. \end{aligned}$$

2. WSPÓLZALEŻNOŚĆ POMIĘDZY BOGACTWEM A INTELIGENCJĄ W SKALI MAKROEKONOMICZNEJ

Dla zilustrowania trudności z jakimi można się spotkać przy analizie współzależności przytoczono wyniki badań R. Lynna oraz T. Vanhanena (2002) nad bogactwem narodów. Przyczyny bogactwa i ubóstwa są przedmiotem zainteresowania wielu badaczy (Landes, 1998), począwszy od pracy Adama Smitha z 1776 roku *Badania nad naturą i przyczynami bogactwa narodów*, poprzez teorie wskazujące na takie determinanty bogactwa narodów jak: klimat (Monteskiusz, 1748; Landes, 2000), kultura i religia (Weber, 1994), położenie geograficzne (Diamond, 1998), system gospodarczy (Olson, 1996), rozwój technologii (Solow, 19956), kapitał ludzki i inteligencja (Herrnstein i Murray, 1996; Lynn i Vanhanen, 2002).

Zależność pomiędzy bogactwem a inteligencją można badać w skali mikroekonomicznej, gdzie jednostkami analizy są osoby. W pracy R. J. Herrnste-

ina i Ch. Murraya (1994) wskazywano, że inteligencja jest dobrym predyktorem sukcesu zawodowego, a niski poziom IQ jest jedną z głównych przyczyn plag społecznych — bezrobocia, przestępczości, narkomanii, samotnego macierzyństwa. Drugim możliwym ujęciem jest analiza makroekonomiczna, w której jednostkami analizy są regiony lub kraje. W pracy R. Lynna i T. Vanhanena (2002) postawiono tezę, że zamożność w skali makroekonomicznej zależy od inteligencji narodu i jest z nią wysoko oraz dodatnio skorelowana.

Zamożność krajów opisano wielkością produktu narodowego brutto (GDP) na osobę w dolarach na podstawie danych dla roku 1998. Analizę oparto na informacjach odnoszących się do 81 krajów, dla których udało się zebrać w miarę porównywalne dane (główne chodziło o trudności w zebraniu odpowiednio dużej liczby informacji opisujących stopień inteligencji). Cecha GDP jest często wykorzystywana w wielu analizach i pomimo pewnych zastrzeżeń co do adekwatności miernika GDP jako dobrego wskaźnika bogactwa na ogół jest akceptowana w wielu opracowaniach.

gorzej przedstawia się sprawa pomiaru inteligencji w skali makroekonomicznej (dla mieszkańców danego kraju). R. Lynn i T. Vanhanen założyli, że inteligencja nie jest równoznaczna z wykształceniem, wiedzą, wynalazczością czy też mądrością i w swoich badaniach mierzyli ją średnim wynikiem testu inteligencji (iloraz inteligencji — IQ) z wielu indywidualnych pomiarów wy-

Tabela 1

Współczynniki inteligencji (IQ), produkt narodowy brutto GDP (w \$ na osobę) oraz liczba ludności (w tys.) w wybranych krajach (dane za różne lata w okresie 1996–2002)

Kraj	IQ	GDP/os. (\$)	Ludność (tys)	Kraj	IQ	GDP/os. (\$)	Ludność (tys)
Argentyna	96	12 013	38 337	Lebanon	86	4 326	3 731
Australia	98	22 452	19 767	Malaysia	92	8 137	23 105
Austria	102	23 166	8 163	Marshall Islands	84	3 000	56
Barbados	78	12 001	277	Mexico	87	7 704	103 718
Belgium	100	23 223	10 331	Morocco	85	3 305	31 754
Brazil	87	6 625	183 960	Nepal	78	1 157	26 470
Bulgaria	93	4 809	7 588	Netherlands	102	22 176	16 223
Canada	97	23 582	31 889	NewZealand	100	17 288	3 960
China	100	3 105	1 291 496	Nigeria	67	795	131 728
Colombia	89	6 006	41 802	Norway	98	26 342	4 555
Congo (Brazz)	73	995	3 413	Peru	90	4 282	27 275
Congo (Zaire)	65	822	56 886	Philippines	86	3 555	86 752

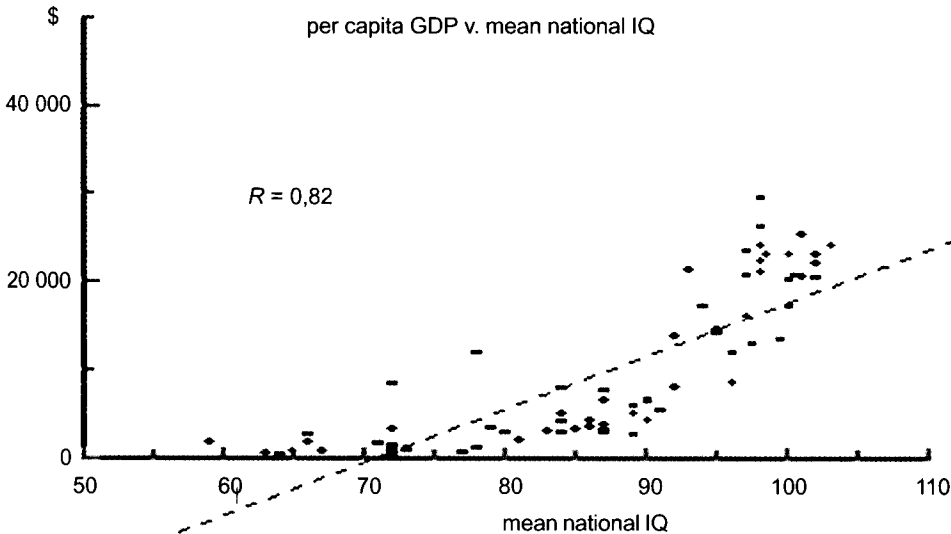
Kraj	IQ	GDP/os. (\$)	Ludność (tys)	Kraj	IQ	GDP/os. (\$)	Ludność (tys)
Croatia	90	6 749	4 498	Poland	99	7 619	38 603
Cuba	85	3 967	11 247	Portugal	95	14 701	10 480
Czech Republic	97	12 362	10 251	Puerto Rico	84	8 000	3 878
Denmark	98	24 218	5 394	Qatar	78	20 987	725
Ecuador	80	3 003	13 074	Romania	94	5 648	22 380
Egypt	83	3 041	74 761	Russia	96	6 460	144 308
Equatorial Guinea	59	1 817	536	Samoa (Western)	87	3 832	204
Ethiopia	63	574	70 366	Sierra Leone	64	458	5 571
Fiji	84	4 231	869	Singapore	103	24 210	4 277
Finland	97	20 847	5 204	Slovakia	96	9 699	5 416
France	98	21 175	62 171	Slovenia	95	14 293	2 012
Germany	102	22 169	82 398	South Africa	72	8 488	46 567
Ghana	71	1 735	21 111	Spain	97	16 212	40 217
Greece	92	13 943	10 626	Sudan	72	1 394	36 593
Guatemala	79	3 505	11 725	Suriname	89	5 161	450
Guinea	66	1 782	8 756	Sweden	101	20 659	8 970
Hong Kong	107	20 763	6 810	Switzerland	101	25 512	7 408
Hungary	99	10 232	10 058	Taiwan	104	13 000	22 543
India	81	2 077	1 057 504	Tanzania	72	480	36 199
Indonesia	89	2 651	223 070	Thailand	91	5 456	63 271
Iran	84	5 121	63 988	Tonga	87	3 000	108
Iraq	87	3 197	24 683	Turkey	90	6 422	68 109
Ireland	93	21 482	3 924	U. Kingdom	100	20 336	60 095
Israel	94	17 301	6 492	Uganda	73	1 074	26 322
Italy	102	20 585	57 998	United States	98	29 605	290 343
Jamaica	72	3 389	2 689	Uruguay	96	8 623	3 387
Japan	105	23 257	127 358	Zambia	77	719	10 800
Kenya	72	980	33 042	Zimbabwe	66	2 669	11 816
Korea, South	106	13 478	47 657				

Źródło: kol. 2-3 — Lynn i Vanhanen, 2002, kol. 4 — <http://www.census.gov/ipc/www/idb/tables.html>.

konanych na mieszkańcach poszczególnych krajów. W sposób bezpośredni udało się im zebrać informacje z 81 krajów (zapewne o różnym stopniu reprezentatywności). W niektórych krajach (Polska, Holandia, Belgia) dane te były gromadzone w ramach widowisk telewizyjnych w stylu: *Narodowy test inteligencji Polaków* (TVN, 2003, 300 tys. osób). R. Lynn i T. Vanhanen oszacowali także średnie ilorazy inteligencji dla pozostałych 104 krajów (dla których dysponowali informacjami o dochodzie narodowym) na podstawie średnich ilorazów inteligencji w sąsiednich krajach, jednakże dane te nie zostały szerzej wykorzystane.

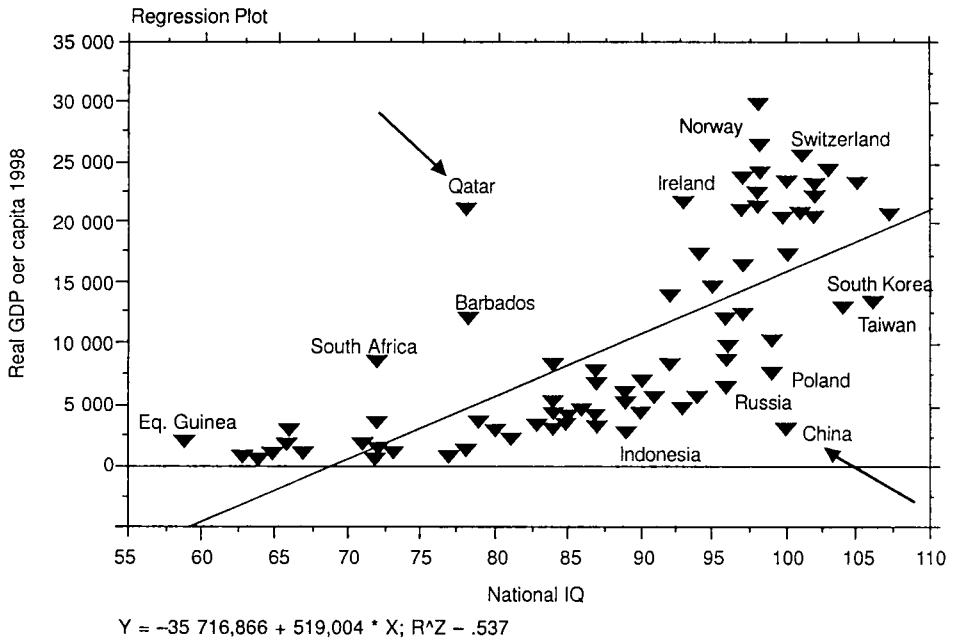
W tabeli 1 przytoczono wartości zmiennych stanowiących podstawę badań R. Lynna i T. Vanhanena. W ostatniej kolumnie przytoczono dodatkowo liczbę mieszkańców dla poszczególnych krajów, gdyż zadaniem niniejszego artykułu jest pokazanie jak się mogą zmienić wyniki analiz w przypadku uwzględnienia dodatkowych informacji (w tym przypadku liczby ludności) o badanych obiektach.

Na rycinach 1–2 przedstawiono rezultaty analizy R. Lynna i T. Vanhanena. Są to diagramy rozrzutu pokazujące zależność wyrażającą się w tym, że kraje zamożne zamieszkałe są na ogół przez mieszkańców z większymi ilorazami inteligencji. I odwrotnie, społeczeństwa krajów niezamożnych zazwyczaj charakteryzują się mniejszymi ilorazami inteligencji. Rycina 1 zawiera oryginalny diagram z pracy R. Lynna i T. Vanhanena natomiast rycina 2 — wykres zmodyfikowany przytoczony w <http://www.amren.com/ar/2002/12/>.



Źródło: <http://www.lagriffedulion.f2s.com/sft.htm>.

Ryc. 1. Oryginalny diagram rozrzutu pomiędzy GDP oraz IQ wraz z liniową funkcją regresji



Źródło: <http://www.amren.com/ar/2002/12/> *The Global Bell Curve. Uncommon sense on wealth and poverty*, reviewed by Thomas Jackson

Ryc. 2. Poprawiony diagram rozrzutu pomiędzy GDP oraz IQ dla 81 krajów wraz z liniową funkcją regresji

Różnica pomiędzy tymi rycinami wyraża się w tym, że rycina 2 odpowiada dokładnie danym zamieszczonym w tabeli 1, natomiast w rycinie 1 nie uwzględniono kilku krajów, m.in. Chin i Kataru, dla których wartości badanych zmiennych wyraźnie odbiegały od linowej linii regresji. Dlatego też współczynniki korelacji pomiędzy GDP a IQ przytoczone w obydwóch rysunkach różnią się od siebie. Na rycinie 1 (bez Chin i Kataru) współczynnik korelacji przyjmuje wartość 0,82, natomiast na rycinie 2 (z Chinami i Kataru) — tylko 0,73. Trudno się oprzeć wrażeniu, że ta operacja mogła być spowodowana chęcią wykazania większej siły zależności. Taki proceder prowokuje do pytania: czy przypadkiem nie dokonywano innych manipulacji danych dla potwierdzenia tezy o wysokim poziomie współzależności pomiędzy inteligencją a bogactwem?

Zależność widoczną na rycinie 2 opisano za pomocą liniowej funkcji regresji o postaci:

$$GDP = -35\,717 + 519 * IQ \quad (R^2 = 0,537).$$

Parametry tej funkcji można zinterpretować następująco. Parametr znajdujący się przy zmiennej IQ (519 \$ GDP/os/rok) wskazuje na „wartość” jednego punktu ilorazu inteligencji (wartość kapitału intelektualnego). Przy wzroście

średniego ilorazu inteligencji mieszkańców danego kraju o 1 punkt IQ można się spodziewać wzrostu zamożności o 519 \$ na 1 osobę w ciągu 1 roku. Jako że inteligencję „produkuje” system oświatowy kraju (szkolnictwo) można stąd oszacować wartość produktu tworzonoego przez ten system, np. w Polsce podniesienie średniego ilorazu inteligencji o 1 punkt IQ może skutkować wzrostem GDP o:

$$[519 \$ * 38 \text{ mln mieszkańców}] \approx 20 \text{ mld } \$/\text{rok} \approx 60 \text{ mld } \text{zł}/\text{rok}.$$

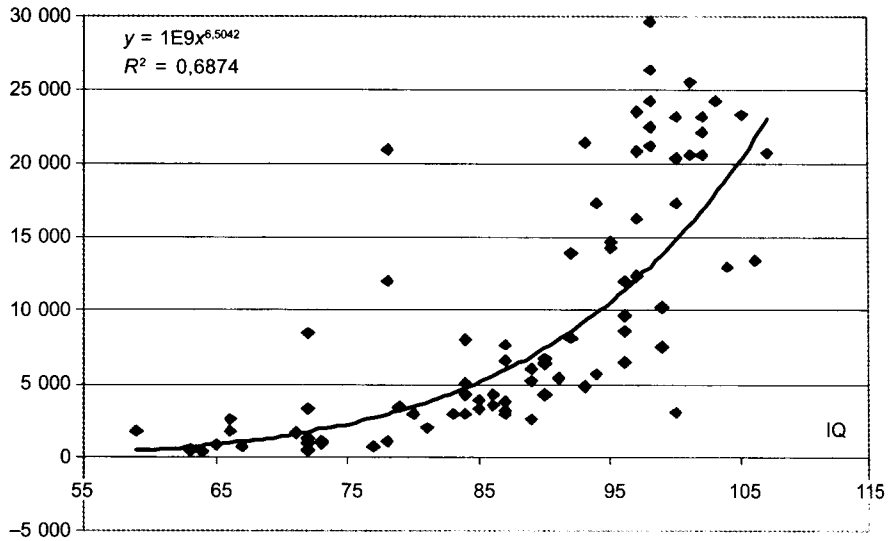
Oczywiście jest to rachunek obarczony znacznym błędem, którego wielkość można oszacować na poziomie $1 - 0,537 = 0,473$, czyli na ok. 47%. Jest zrozumiałe, że wzrost produktu narodowego zależy od wielu innych czynników, a poziom inteligencji (o ile w ogóle) jest tylko jednym z nich. Tym niemniej, nawet uwzględniając ten wysoki brak precyzji, sektory oświaty, szkolnictwa wyższego i nauki mają pewne podstawy do twierdzenia, że wartość ich działalności (po udowodnieniu, że iloraz inteligencji społeczeństwa w ciągu roku rzeczywiście wzrósł o 1 pkt IQ) kształtuje się co najmniej na poziomie ok. 30 mld zł w skali roku (tzn. 50% z 60 mld zł).

Interpretacja wyrazu wolnego jest bardziej abstrakcyjna. Formalnie jest to wielkość GDP w przypadku „zerowej” inteligencji mieszkańców danego kraju. W tym przypadku oznacza to, że do każdego „mieszkańca” należałoby dolożyć ponad 35 tys. \$ GDP na rok (ujemna wartość wyrazu wolnego funkcji regresji) aby społeczeństwo takie mogło w ogóle funkcjonować.

Bliższa analiza rycin 1 i 2 wskazuje, że liniowa funkcja regresji nie oddaje zbyt poprawnie empirycznego rozrzutu punktów. W szczególności dla krajów o niskich ilorazach inteligencji teoretyczne wielkości GDP są ujemne, podczas gdy faktyczne wielkości w tych krajach są dodatnie. Dlatego też na podstawie tych samych danych oszacowano dwie inne funkcje regresji: potęgową oraz wykładniczą. Wyniki w postaci diagramów rozrzutu z dopasowanymi funkcjami regresji przedstawiono na rycinach 3 i 4.

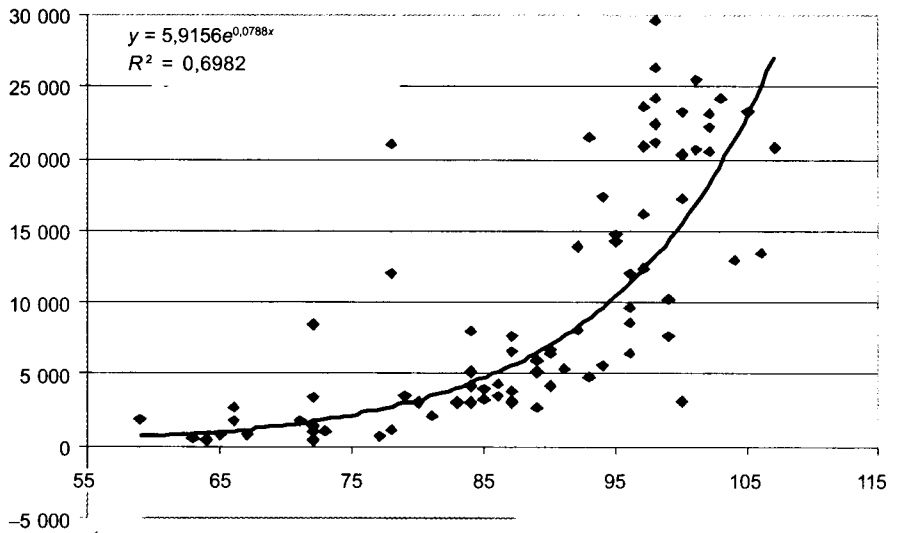
Jak można zauważyć obydwie funkcje są lepiej dopasowane do danych empirycznych, jakkolwiek funkcja wykładnicza jest nieco „lepsza” ($R^2 = 0,698$) od funkcji potęgowej ($R^2 = 0,687$). W przypadku funkcji liniowej współczynnik determinacji był mniejszy i wynosił tylko $R^2 = 0,537$. Parametry oszacowanych funkcji regresji zebrano w tabeli 2. Obok parametrów strukturalnych podano tu odpowiadające im statystyki Studenta, miary dopasowania (współczynniki korelacji i determinacji, średni błąd dopasowania i współczynnik zmienności resztowej), jak również teoretyczne wielkości GDP dla Polski wynikające z oszacowanych funkcji regresji oraz względne wielkości różnicy pomiędzy teoretycznymi a faktyczną (7619 \$/os/rok) wielkością GDP dla Polski.

Ten ostatni parametr można zinterpretować jako tę wielkość produktu narodowego brutto, która powinna uzyskać Polska w relacji do średniego ilorazu inteligencji IQ mieszkańców naszego kraju (oczywiście w sytuacji, gdyby



Źródło: obliczenia własne.

Ryc. 3. Potęgowa funkcja regresji pomiędzy GDP a IQ (81 krajów)



Źródło: obliczenia własne.

Ryc. 4. Wykładnicza funkcja regresji pomiędzy GDP a IQ (81 krajów)

inteligencja była jedynym czynnikiem determinującym poziom GDP). Jak się okazuje, faktyczna wielkość GDP (7619 \$/os/rok) jest o połowę mniejsza niż wielkości teoretyczne wynikające z oszacowanych funkcji regresji. Można to traktować jako szacunek wielkości niewykorzystanego ekonomicznie potencjału tkwiącego w kapitale ludzkim społeczeństwa Polski.

Tabela 2

Parametry liniowej, potęgowej i wykładniczej funkcji regresji pomiędzy GDP a IQ

Parametr	Funkcja liniowa	Funkcja potęgowa	Funkcja wykładnicza
Wyraz wolny [a]	-35 717	1,4545E-09	5,9156
Statystyka Studenta [t(a)]	-7,40	-9,22	3,42
Współczynnik kierunkowy [b]	519,0	6,504	0,0788
Statystyka Studenta [t(b)]	9,57	13,18	13,52
Współczynnik determinacji [R ²]	0,537	0,687	0,698
Współczynnik korelacji [R]	0,732	0,829	0,836
Średni błąd dopasowania [s]	5 813 \$	5 672 \$	5 665 \$
Współczynnik. zmienności resztowej [V]	57,8%	56,4%	56,3%
Wartość teoretyczna GDP dla Polski	15 664 \$	13 893 \$	14 391 \$
Względna różnica między teoretycznym a faktycznym GDP dla Polski	-51,4%	-45,2%	-47,1%

Źródło: obliczenia własne.

Oczywiście interpretacja tych parametrów jest dyskusyjna i trudno ją rekomendować. Pokazuje tylko formalne możliwości wykorzystania klasycznych modeli regresji.

3. MODYFIKACJE ANALIZOWANYCH ZMIENNYCH

W przedstawionej powyżej analizie wykorzystano standardowe podejście, w którym rozpatruje się zależności korelacyjno-regresyjne. Na pierwszy rzut oka jednak widać, że dyskusyjne jest identyczne traktowanie w ocenie współzależności takich krajów jak Chiny (1,3 mld mieszkańców, czyli 1/4 mieszkańców całej badanej zbiorowości) oraz np. Wysp Marshalla (50 tys. mieszkańców) czy też Wysp Tonga (100 tys. mieszkańców), które to państwa mogłyby być traktowane jako niewielkie miasta. Tymczasem na diagramach rozrzutu czy też w obliczeniach wszystkie obiekty (kraje) mają takie samo znaczenie (wagę).

W tabeli 3 podano skumulowane wartości (w procentach) trzech zmiennych — liczby ludności [LUD] oraz iloczynów liczby ludności przez średnie ilorazy inteligencji [IQ*LUD] a także przez wskaźniki produktu narodowego brutto [GDP*LUD]. Kraje uporządkowane są według rosnącej liczby ludności,

Skumulowana liczba ludności (LUD), kapitału intelektualnego (IQ*LUD)
oraz kapitału ekonomicznego (GDP*LUD)

Lp.	Kraj	LUD (tys.)	IQ*LUD/ 100	GDP*LUD/ 1000	% CUM LUD	% CUM (IQ*LUD)	% CUM (GDP*LUD)
1	Marshall Islands	56	47	169	0,00	0,00	0,00
2	Tonga	108	94	325	0,00	0,00	0,00
3	Samoa (Western)	204	177	781	0,01	0,01	0,00
4	Barbados	277	216	3 321	0,01	0,01	0,03
5	Suriname	450	400	2 322	0,02	0,03	0,01
6	Equatorial Guinea	536	316	974	0,03	0,02	0,01
7	Qatar	725	566	15 218	0,05	0,04	0,14
8	Fiji	869	730	3 676	0,06	0,05	0,05
9	Slovenia	2 012	1 911	28 752	0,10	0,10	0,42
10	Jamaica	2 689	1 936	9 114	0,15	0,14	0,09
11	Uruguay	3 387	3 252	29 209	0,22	0,33	0,50
12	Congo (Brazz)	3 413	2 491	3 396	0,28	0,19	0,04
13	Lebanon	3 731	3 209	16 140	0,36	0,26	0,22
14	Puerto Rico	3 878	3 257	31 023	0,43	0,40	0,75
15	Ireland	3 924	3 649	84 296	0,51	0,55	2,62
16	New Zealand	3 960	3 960	68 465	0,58	0,64	2,17
17	Singapore	4 277	4 405	103 541	0,67	0,82	3,18
18	Croatia	4 498	4 048	30 356	0,75	0,72	0,59
19	Norway	4 555	4 464	119 998	0,84	0,91	5,00
20	Finland	5 204	5 048	108 496	0,94	1,02	4,05
21	Denmark	5 394	5 286	130 635	1,04	1,25	6,04
22	Slovakia	5 416	5 200	52 534	1,15	1,13	1,99
23	Sierra Leone	5 571	3 565	2 551	1,26	0,47	0,02
24	Israel	6 492	6 102	112 317	1,38	1,50	4,35
25	Hong Kong	6 810	7 286	141 391	1,51	1,81	6,43
26	Switzerland	7 408	7 482	189 001	1,66	1,97	8,78
27	Bulgaria	7 588	7 057	36 493	1,80	1,65	1,03
28	Austria	8 163	8 326	189 096	1,96	2,49	9,30
29	Guinea	8 756	5 779	15 602	2,13	1,37	0,18
30	Sweden	8 970	9 060	185 318	2,30	2,68	7,76

Lp.	Kraj	LUD (tys.)	IQ*LUD/ 100	GDP*LUD/ 1000	% CUM LUD	% CUM (IQ*LUD)	% CUM (GDP*LUD)
31	Hungary	10 058	9 957	102 911	2,50	3,93	2,90
32	Czech Republic	10 251	9 944	126 724	2,69	3,51	5,69
33	Belgium	10 331	10 331	239 913	2,89	4,16	10,57
34	Portugal	10 480	9 956	154 066	3,10	3,72	7,25
35	Greece	10 626	9 776	148 158	3,30	3,30	6,83
36	Zambia	10 800	8 316	7 765	3,51	2,31	0,07
37	Cuba	11 247	9 560	44 616	3,73	3,09	1,58
38	Guatemala	11 725	9 263	41 097	3,95	2,88	1,46
39	Zimbabwe	11 816	7 798	31 536	4,18	2,13	0,84
40	Ecuador	13 074	10 459	39 261	4,43	4,38	1,23
41	Netherlands	16 223	16 548	359 767	4,75	5,05	16,51
42	Australia	19 767	19 371	443 798	5,13	5,88	19,99
43	Ghana	21 111	14 989	36 628	5,53	4,70	1,13
44	Romania	22 380	21 037	126 404	5,97	6,77	5,34
45	Taiwan	22 543	23 445	293 065	6,40	8,19	12,06
46	Malaysia	23 105	21 257	188 006	6,85	7,23	8,27
47	Iraq	24 683	21 474	78 913	7,32	7,69	2,39
48	Uganda	26 322	19 215	28 270	7,83	5,47	0,35
49	Nepal	26 470	20 646	30 625	8,34	6,32	0,67
50	Peru	27 275	24 548	116 792	8,87	9,22	4,67
51	Morocco	31 754	26 991	104 948	9,48	10,92	3,75
52	Canada	31 889	30 933	752 014	10,10	11,58	28,43
53	Kenya	33 042	23 790	32 381	10,73	8,70	0,93
54	Tanzania	36 199	26 063	17 375	11,43	9,78	0,27
55	Sudan	36 593	26 347	51 010	12,14	10,34	1,84
56	Argentina	38 337	36 803	460 541	12,88	13,09	21,24
57	Poland	38 603	38 217	294 115	13,62	15,49	12,86
58	Spain	40 217	39 011	652 005	14,40	16,33	26,38
59	Colombia	41 802	37 204	251 062	15,21	14,68	11,26
60	South Africa	46 567	33 528	395 258	16,10	12,30	17,59
61	Korea, South	47 657	50 516	642 316	17,02	18,36	24,60
62	Congo (Zaire)	56 886	36 976	46 760	18,12	13,88	1,71

Lp.	Kraj	LUD (tys.)	IQ*LUD/ 100	GDP*LUD/ 1000	% CUM LUD	% CUM (IQ*LUD)	% CUM (GDP*LUD)
63	Italy	57 998	59 158	1 193 896	19,24	22,01	36,40
64	U. Kingdom	60 095	60 095	1 222 085	20,40	23,29	43,04
65	France	62 171	60 928	1 316 479	21,60	24,60	46,63
66	Thailand	63 271	57 577	345 207	22,82	20,74	15,53
67	Iran	63 988	53 750	327 684	24,06	19,51	14,59
68	Turkey	68 109	61 299	437 399	25,37	25,91	18,78
69	Ethiopia	70 366	44 331	40 390	26,73	17,28	1,34
70	Egypt	74 761	62 051	227 347	28,17	27,23	9,92
71	Germany	82 398	84 046	1 826 688	29,76	30,63	51,61
72	Philippines	86 752	74 606	308 402	31,43	28,83	13,70
73	Mexico	103 718	90 235	799 044	33,43	34,45	30,60
74	Japan	127 358	133 726	2 961 959	35,89	37,31	65,66
75	Nigeria	131 728	88 257	104 723	38,43	32,52	3,47
76	Russia	144 308	138 536	932 229	41,22	40,27	33,14
77	Brazil	183 960	160 045	1 218 734	44,77	43,70	39,72
78	Indonesia	223 070	198 532	591 358	49,07	47,95	22,85
79	United States	290 343	284 536	8 595 591	54,67	54,04	100,00
80	India	1 057 504	856 579	2 196 437	75,08	72,36	57,59
81	China	1 291 496	1 291 496	4 010 095	100,00	100,00	76,58

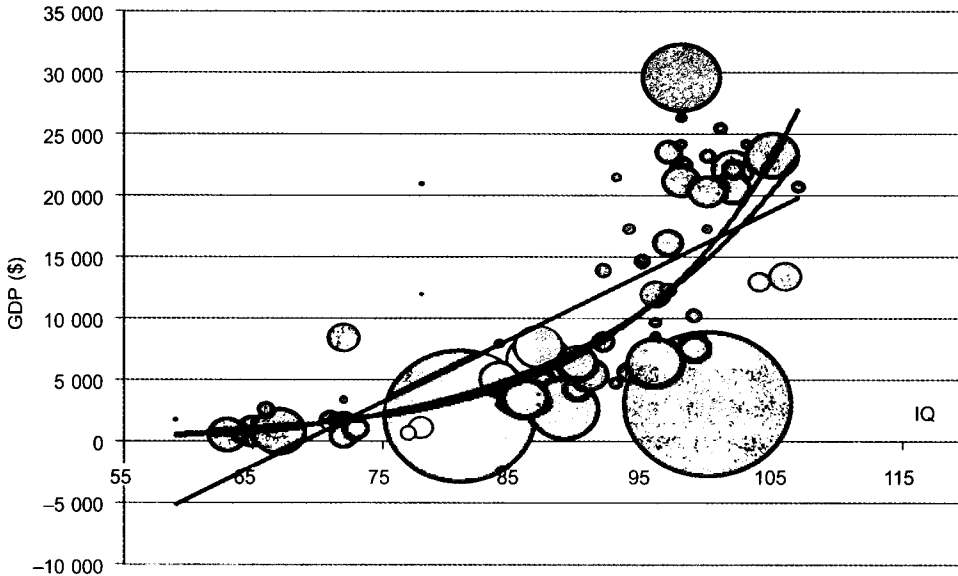
Źródło: obliczenia własne.

wskutek czego kumulanty zmiennych [IQ*LUD] i [GDP*LUD] wskazują na inną kolejność krajów. Zmienną [IQ*LUD] można interpretować jako całkowity zasób kapitału intelektualnego znajdującego się w dyspozycji danego kraju, natomiast zmienną [GDP*LUD] jako całkowity zasób produktu narodowy brutto w tym kraju.

Jak można zauważyć z tabeli 3 łączna liczba mieszkańców połowy (50%) najmniejszych krajów wyczerpuje zaledwie 5% populacji wszystkich analizowanych krajów. Podobnie jest z zasobem kapitału intelektualnego [IQ*LUD]. Z drugiej strony liczba mieszkańców samych tylko Chin odpowiada skumulowanej liczbie mieszkańców 67 krajów spośród 81, a zasób kapitału intelektualnego Chin jest taki jak kapitał 70 mniejszych krajów. Liczba mieszkańców i kapitał intelektualny dwóch największych krajów — Chin i Indii odpowiada prawie połowie zasobów tych zmiennych w całym analizowanym zbiorze krajów.

Kumulanty liczby ludności i kapitału intelektualnego praktycznie się pokrywają ($r = 0,99$) natomiast kumulanta całkowitego produktu narodowego [GDP*LUD] kształtuje się nieco inaczej niż poprzednie dwie kumulanty. Wynika to stąd, że największe udziały w zakresie [GDP*LUD] mają Stany Zjednoczone (30% ogólnej sumy), a nie Chiny (14,2%) czy Indie (7,7%). Ponadto dochodzi tu jeszcze Japonia (10% potencjału gospodarczego wobec 2,5% udziału w liczbie ludności oraz 4% udziału w zasobie kapitału intelektualnego).

Zróznicowane wielkości krajów powodują, że uzasadnionym wydaje się zaproponowanie zmodyfikowanej formuły wyznaczania miar zależności. Na rycinie 5 przedstawiono bąbelkowy diagram rozrzutu wraz z wszystkimi trzema funkcjami regresji. Średnica poszczególnych okręgów(krajów) jest tu proporcjonalna do liczby ludności [LUD]. Wykres ten lepiej oddaje istotę rozpatrywanej zależności i sugeruje uwzględnienie wielkości obiektów w trakcie wyznaczania miar korelacji i regresji.



Źródło: obliczenia własne.

Ryc. 5. Bąbelkowy diagram rozrzutu pomiędzy GDP a IQ wraz z funkcjami regresji

4. PROPOZYCJE ZMODYFIKOWANYCH MIERNKÓW WSPÓLZALEŻNOŚCI

Zgodnie z dotychczas stosowanym podejściem w analizie zależności w większości badań empirycznych współczynnik korelacji liniowej wyznaczany jest według następującego wzoru:

$$r = \frac{\text{cov}(X, Y)}{s_X \cdot s_Y} = \frac{\sum_{i=1}^n X_i Y_i - \frac{\left(\sum_{i=1}^n X_i\right) \left(\sum_{i=1}^n Y_i\right)}{n}}{\sqrt{\left(\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}\right) \left(\sum_{i=1}^n Y_i^2 - \frac{\left(\sum_{i=1}^n Y_i\right)^2}{n}\right)}}, \quad (1)$$

gdzie $\text{cov}(X, Y)$ to kowariancja pomiędzy zmiennymi X i Y natomiast s_X, s_Y to ich odchylenia standardowe.

W omawianym przykładzie zmienną Y jest wskaźnik produktu narodowego brutto (GDP), natomiast zmienną X — średni iloraz inteligencji (IQ). Wzór (1) można przekształcić w ten sposób, że przy każdym jego elemencie pojawi się waga F_i (równa liczbie ludności dla i -tego kraju — LUD).

$$r = \frac{\sum_{i=1}^n F_i X_i Y_i - \frac{\left(\sum_{i=1}^n F_i X_i\right) \left(\sum_{i=1}^n F_i Y_i\right)}{\sum_{i=1}^n F_i}}{\sqrt{\left(\sum_{i=1}^n F_i X_i^2 - \frac{\left(\sum_{i=1}^n F_i X_i\right)^2}{\sum_{i=1}^n F_i}\right) \left(\sum_{i=1}^n F_i Y_i^2 - \frac{\left(\sum_{i=1}^n F_i Y_i\right)^2}{\sum_{i=1}^n F_i}\right)}}. \quad (2)$$

Aby przybliżyć interpretację współczynnika danego wzorem (2) należy zauważyć, że mierzy on zależność pomiędzy kapitałem intelektualnym a potencjałem gospodarczym w sytuacji, w której jednostkami obserwacji są wszystkie osoby (a nie kraje) zamieszkałe w analizowanych 81 krajach, przy czym wszyscy mieszkańcy danego kraju mają jednakowy (na średnim poziomie) iloraz inteligencji oraz są twórcami jednakowego (na średnim poziomie) produktu narodowego brutto.

Inną propozycją jest wyznaczenie współczynnika korelacji liniowej pomiędzy zmiennymi $[IQ \cdot LUD]$ oraz $[GDP \cdot LUD]$, co sprowadza się do następującej formuły:

$$r = \frac{\sum_{i=1}^n (F_i X_i)(F_i Y_i) - \frac{\left(\sum_{i=1}^n (F_i X_i)\right) \left(\sum_{i=1}^n (F_i Y_i)\right)}{n}}{\sqrt{\left(\sum_{i=1}^n (F_i X_i)^2 - \frac{\left(\sum_{i=1}^n (F_i X_i)\right)^2}{n}\right) \left(\sum_{i=1}^n (F_i Y_i)^2 - \frac{\left(\sum_{i=1}^n (F_i Y_i)\right)^2}{n}\right)}}, \quad (3)$$

względnie do formuły zmodyfikowanej, w której przez liczbę obserwacji rozumie się sumę wszystkich realizacji zmiennej wykorzystywanej w charakterze wag F :

$$r = \frac{\sum_{i=1}^n (F_i X_i)(F_i Y_i) - \frac{\left(\sum_{i=1}^n (F_i X_i)\right)\left(\sum_{i=1}^n (F_i Y_i)\right)}{\sum_{i=1}^n F_i}}{\sqrt{\left(\sum_{i=1}^n (F_i X_i)^2 - \frac{\left(\sum_{i=1}^n (F_i X_i)\right)^2}{\sum_{i=1}^n F_i}\right)\left(\sum_{i=1}^n (F_i Y_i)^2 - \frac{\left(\sum_{i=1}^n (F_i Y_i)\right)^2}{\sum_{i=1}^n F_i}\right)}} \quad (4)$$

Analogicznie do przedstawionych powyżej wzorów można podać cztery różne formuły służące do wyznaczania parametrów liniowej funkcji regresji $Y = f(X)$. Zebrano je w tabeli 4.

Tabela 4

Wzory na nieważone (1) i ważone (2)–(4) parametry liniowej funkcji regresji

Wzór	Współczynnik kierunkowy	Wyraz wolny
(1)	$b = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2}$	$a = \frac{\sum_{i=1}^n Y_i}{n} - b \frac{\sum_{i=1}^n X_i}{n}$
(2)	$b = \frac{\sum_{i=1}^n F_i \sum_{i=1}^n F_i X_i Y_i - \sum_{i=1}^n F_i X_i \sum_{i=1}^n F_i Y_i}{\sum_{i=1}^n F_i \sum_{i=1}^n F_i X_i^2 - \left(\sum_{i=1}^n F_i X_i\right)^2}$	$a = \frac{\sum_{i=1}^n (F_i Y_i)}{\sum_{i=1}^n F_i} - b \frac{\sum_{i=1}^n (F_i X_i)}{\sum_{i=1}^n F_i}$
(3)	$b = \frac{n \sum_{i=1}^n (F_i X_i)(F_i Y_i) - \sum_{i=1}^n (F_i X_i) \sum_{i=1}^n (F_i Y_i)}{n \sum_{i=1}^n (F_i X_i)^2 - \left(\sum_{i=1}^n (F_i X_i)\right)^2}$	$a = \frac{\sum_{i=1}^n (F_i Y_i)}{n} - b \frac{\sum_{i=1}^n (F_i X_i)}{n}$
(4)	$b = \frac{\sum_{i=1}^n F_i \sum_{i=1}^n (F_i X_i)(F_i Y_i) - \sum_{i=1}^n (F_i X_i) \sum_{i=1}^n (F_i Y_i)}{\sum_{i=1}^n F_i \sum_{i=1}^n (F_i X_i)^2 - \left(\sum_{i=1}^n (F_i X_i)\right)^2}$	$a = \frac{\sum_{i=1}^n (F_i Y_i)}{\sum_{i=1}^n F_i} - b \frac{\sum_{i=1}^n (F_i X_i)}{\sum_{i=1}^n F_i}$

Źródło: obliczenia własne.

Wartości współczynników regresji oraz korelacji liniowej wyznaczone za pomocą powyższych wzorów przedstawiono w tabeli 5. Jak można zauważyć ważne miary zależności (2) dają nieco inne rezultaty aniżeli standardowe miary nieważone (1). Współczynnik korelacji zmniejszył się z 0,73 do 0,48. Również zmniejszyły się parametry linowej funkcji regresji — współczynnik kierunkowy z 519 \$ GDP/os/rok do 369 \$ GDP/os/rok i podobnie wyraz wolny, z -35717 \$ GDP do -26216 \$ GDP.

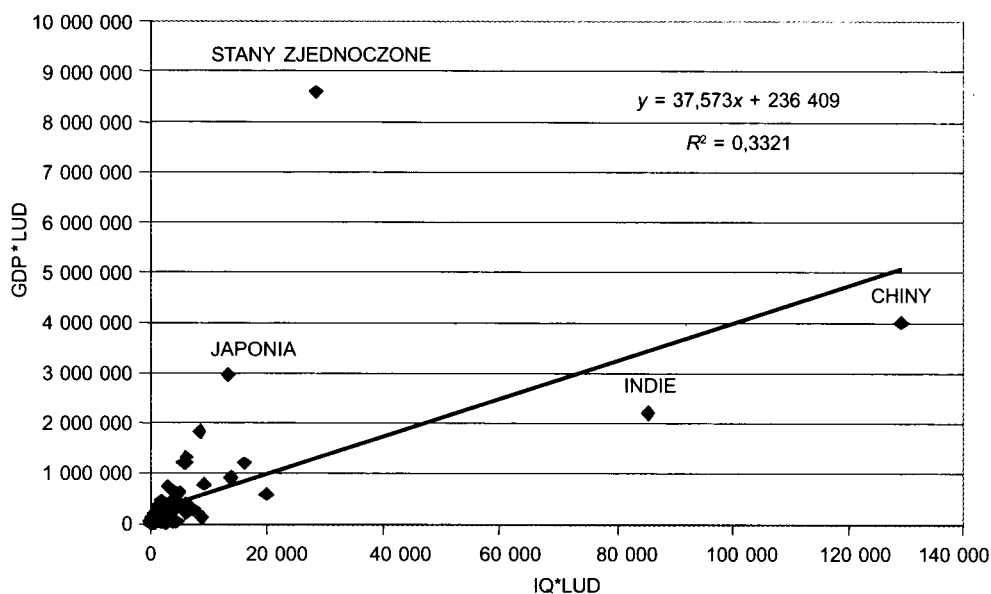
Tabela 5

Ważone parametry regresji i współczynniki korelacji pomiędzy GDP a IQ

Wzór	Współczynnik a	Współczynnik b	Współczynnik r
(1)	-35 717	519	0,733
(2)	-26 215	369	0,481
(3)	236 409	38	0,576
(4)	3 326	42	0,625

Źródło: obliczenia własne.

Miary zależności opisane za pomocą trzeciej i czwartej formuły dają odmiennie rezultaty. Jak wynika z ryciny 6 kształt diagramu rozrzutu dla zmiennych [IQ*LUD] oraz [GDP*LUD] determinują 3–4 kraje: Chiny, Stany Zjedno-



Źródło: obliczenia własne.

Ryc. 6. Diagram rozrzutu wraz z liniową funkcją regresji pomiędzy zmiennymi GDP*LUD oraz IQ*LUD

zione, Indie oraz Japonia i one decydują o ocenie kierunku oraz sile zależności pomiędzy potencjałem intelektualnym a bogactwem. Pozostałe kraje tworzą jedno skupisko zgrupowane w pobliżu początku układu współrzędnych (0, 0). „Wartość” jednostkowa kapitału intelektualnego spada tu z 369–519 \$ do 38–42 \$ GDP na 1 osobę w ciągu roku.

BIBLIOGRAFIA

- Herrnstein R., Murray Ch. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*, Free Press 1994.
- [http://www.amren.com/ar/2002/12/ The Global Bell Curve. Uncommon sense on wealth and poverty, reviewed by Thomas Jackson.](http://www.amren.com/ar/2002/12/The_Global_Bell_Curve._Uncommon_sense_on_wealth_and_poverty,_reviewed_by_Thomas_Jackson.)
- [http://www.census.gov/ipc/www/idb/tables.html.](http://www.census.gov/ipc/www/idb/tables.html)
- [http://www.lagriffedulion.f2s.com/sft.htm.](http://www.lagriffedulion.f2s.com/sft.htm)
- Human Development Report 2000*. 2000. United Nations Development Programme (UNDP), Oxford University Press, New York.
- Jensen A.R. 1998. *The g Factor*, Praeger, Westport (CT).
- Landes D.S. 1998. *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*, W.W. Norton & Company, New York.
- Lynn R. 2008. *The Global Bell Curve*, Augusta, GA: Washington Summit Publisher, 2008.
- Lynn R., Vanhanen T. *IQ and the Wealth of Nations*. Westport, CT Praeger, 2002.
- Lynn R., Vanhanen T. *IQ and Global Inequality*. Washington Summit Publishers (GA), Augusta.
- Rushton J.P. 2000. *Race, Evolution, and Behavior: A Life History Perspective*, Charles Darwin Research Institute (MI), Port Huron.

MARKOV SWITCHING IN STOCHASTIC VARIANCE. BAYESIAN COMPARISON OF TWO SIMPLE MODELS

ŁUKASZ KWIATKOWSKI¹

Katedra Metod Statystycznych
Krakowska Akademia
im. Andrzeja Frycza Modrzewskiego
PL 30-705 Kraków, ul. Gustawa Herlinga-Grudzińskiego 1
e-mail: kwiatkowski.lukas@gmail.com

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie 13 stycznia 2009 roku przez autora.

ABSTRACT

Ł. Kwiatkowski. *Markov Switching in Stochastic Variance. Bayesian Comparison of Two Simple Models.* Folia Oeconomica Cracoviensia 2008–2009, 49–50: 109–143.

In the paper two particular Markov Switching Stochastic Volatility models (MSSV) are under consideration: one with a switching intercept in the log-volatility equation, and the other — with a regime-dependent autoregression parameter. While the former one is fairly common in the literature (as a tool of taking account for regimes of different mean volatility level), the latter has not been paid almost any attention so far. We note the fact, that state-varying mean volatility may arise from switches in the intercept or in the autoregression parameter. Hence, we aim to compare these two models in respect of goodness of fit to the data from the Polish financial market, employing Bayesian techniques of estimation and model comparison. Clear evidence of structural shifts in the volatility pattern is found. Two different regimes of the economy are characterized in terms of the mean volatility level and the variance of volatility.

KEY WORDS — SŁOWA KLUCZOWE

Markov switching, Stochastic Volatility model, Bayesian inference
przełączanie typu Markowa, model zmienności stochastycznej (SV), wnioskowanie bayesowskie

¹ Assistant at The Andrzej Frycz Modrzewski Krakow University College, Department of Statistics. The author is deeply indebted to Prof. J. Osiewalski and Dr. A. Pajor for their invaluable help and advice.

1. INTRODUCTION

Recent years have witnessed a rapid increase in the interest in modelling time-varying volatility of financial time series. Among the most popular tools devised to capture some of its common features have been two parametric model families: the ARCH processes, introduced by Engle (1982) (along with their numerous extensions starting with GARCH process of Bollerslev, 1986), and the stochastic volatility (SV) processes, of whose the main concept has been presented by Clark (1973). Although in formally different ways, in both of the above conditional variance equation is defined explicitly.

The underlying assumption of these constructions is homogeneity of the modelled time series, which means exclusion of potential structural breaks occurring in the analyzed period. It allows one to presume that the parameters of interest remain constant over time. However, volatility clustering, a common phenomenon observed in stock returns series, may question this belief. It is so due to some heuristic reasoning that less volatile periods alternating with these of higher uncertainty may somehow correspond with structural breaks occurring in the data. In view of potential heterogeneity of a certain time series, models such as GARCH or SV are of too restrictive nature (Hwang et al., 2004). Not being able to capture discrete shifts of states of the economy may be the cause for these models to yield somewhat misleading results. For instance, Granger and Hyung (1999) and Diebold and Inoue (2001) suggest that structural breaks in the mean of volatility may be a source of volatility persistence. It follows that a proper model should include an explicit mechanism capable of accounting for possible regime changes. One of the most popular in this respect is Markov switching (MS) mechanism introduced by Hamilton (1989). What he suggested is an autoregressive process whose parameters are subject to changes over time according to a latent homogeneous Markov chain. Since then many studies have been undertaken to employ the idea of MS into volatility models, mainly those of the GARCH family (see Bauwens et al., 2006, among many).

The aim of the paper is Bayesian estimation and comparison (in terms of goodness of fit to the data) three SV specifications: a non-switching basic stochastic volatility (BSV) model and two Markov Switching SV (MSSV) models (one with a regime-changing intercept and the other with a switching autoregression parameter in the volatility equation). The dataset comprises daily observations on the 1-month Warsaw Interbank Offered Rate (WIBOR1M) interest rates over the period from April 17, 2000 to April 7, 2008. Incidental to the analysis of the regime-switching constructions is a search for potential structural shifts occurring in the series and — if any are found — characterization of the identified states of the economy.

There are several reasons behind our research. Firstly, employing non-switching models in view of potential structural breaks in the time series may

lead to a model misspecification error. In this regard, switching specifications, like the MSSV processes, may be of value as they account for discrete shifts in the parameters. Secondly, we note that abrupt changes in the mean volatility level, which is the reason widely cited in the literature for employing the MSSV structures, may be attributed not only to the switching intercept but, alternatively, to the regime-changing autoregression parameter in the volatility equation. Finally, neither the issue above nor the MSSV models with a switching elasticity of volatility² are tackled in the literature known to the author.³

As regards the current state of the literature on the MSSV models, in a predominant part of the studies only two-state specifications with a switching intercept are concerned (Smith, 2000; Kalimipali and Susmel, 2001; Casarin, 2003; Shibata and Watanabe, 2005; Carvalho and Lopes, 2006). Three-state models are analyzed in So et al. (1998) and Hwang et al. (2003, 2004). In terms of the estimation tools the Bayesian approach prevails, with use of either standard MCMC procedures (the Gibbs sampler; So et al., 1998; Kalimipali and Susmel, 2001, Shibata and Watanabe, 2005) or more recent (auxiliary) particle filters (Casarin, 2003; Carvalho and Lopes, 2006). Some of the models feature additional elements such as term structure (Smith, 2000; Kalimipali and Susmel, 2001) and heavy-tailed distributions of the noise term in the observable process (Casarin, 2003).

We conduct the analysis in the Bayesian setting, which allows fully probabilistic inference on all the unknown quantities of the model as well as well-founded model comparison. As opposed to the 'classical' (i.e. non-Bayesian) tools, Bayesian methodology in the context of switching models (or, more generally, mixture models) is found even more appealing. The latter stems from the possibility of inference on the latent regimes unconditionally upon the parameter estimates (see Gärtner, 2007).

The remainder of the paper is organized as follows. In Section 2 we present the models under consideration and selected regime characteristics, of which use is made in the further parts. Bayesian estimation of the models and their comparison are briefly discussed in Section 3, followed by an empirical illustration of the presented methodology in Section 4. Finally, Section 5 concludes.

² 'Elasticity of volatility' is the term used by Smith (2000) with reference to the autoregression parameter, ϕ , in the log-volatility equation of a SV model given as: $\ln h_t = \mu + \phi \ln h_{t-1} + \sigma \eta_t$. Assuming $\eta_t \sim iid(0, 1)$ (i.e. each η_t is an independent and identically distributed random variable with zero mean and unit variance) the third parameter, σ , is a standard deviation of the innovation term $\sigma \eta_t$, and hence referred to as 'volatility of volatility'.

³ The only works in which the autoregression parameter is allowed to switch over the regimes are of Hwang et al. (2003, 2004). However, not only the estimation approach employed in these studies (Quasi-ML), but also the specification of the log-volatility equation is different than in our work.

2. SELECTED MARKOV SWITCHING SV (MSSV) MODELS

In this part the basics of selected MSSV processes are presented. We start with the following definition of a general Markov Switching SV process.

Definition 1

A stochastic process⁴ $\{y_t, t \in N \cup \{0\}\}$ follows a two-state Markov Switching Stochastic Volatility (MSSV) process if and only if for each $t \in N \cup \{0\}$ the following assumptions hold:

$$y_t = \varepsilon_t \sqrt{h_t}; \quad (1)$$

$$\ln h_t = \mu_{S_t} + \varphi_{S_t} \ln h_{t-1} + \sigma_{S_t} \eta_t; \quad (2)$$

$$\left\{ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \middle| t \in N \cup \{0\} \right\} \sim iiN^{(2)}(\mathbf{0}_{(2 \times 1)}, I_2);$$

$\{S_t, t \in N \cup \{0\}\}$ — a homogenous, ergodic and irreducible two-state Markov chain;

$$S_t \in Q = \{1, 2\};$$

$$\Pr(S_t = j \mid S_{t-1} = i) = p_{ij}, \quad p_{ij} \in (0, 1), \quad \sum_{j=1}^2 p_{ij} = 1, \quad i, j = 1, 2.$$

The observable variable, y_t , is defined as a product of a Gaussian white noise and conditional standard deviation.⁵ Equation 2 defines the log-volatility which evolves over time according to a simple switching autoregressive process. Since all of the parameters in the latter feature regime-changing property, the definition may be viewed quite general, although further extensions are possible (a heavy-tailed distribution for ε_t can be considered, for instance, as in Casarin, 2003). The switching mechanism, represented by the family of discrete random variables S_t 's, is assumed to follow a simple two-state Markov chain, in accordance with the idea proposed by Hamilton (1989). For the sake of our study, ergodicity and irreducibility of the chain are assumed by restricting the transition probabilities, p_{ij} , to lay strictly within the unit interval.

One should note, that a basic stochastic volatility process (BSV), with the log-volatility equation defined as:

$$\ln h_t = \mu + \varphi \ln h_{t-1} + \sigma \eta_t,$$

⁴ By N we denote the set of positive integers.

⁵ It is straightforward to show that — conditionally upon a σ -algebra with respect to which h_t is measurable — the latter constitutes conditional variance of the process $\{y_t\}$, i.e: $\text{Var}(y_t \mid F_{t-1}, \eta_t, S_t) = h_t$, where F_{t-1} is the past information about the process $\{h_t, t \in N \cup \{0\}\}$ up to the moment $t-1$.

may be viewed as a particular case of the general MSSV process once $\mu_1 = \mu_2$, $\varphi_1 = \varphi_2$ and $\sigma_1 = \sigma_2$ hold. However, the transition probabilities, p_{ij} , remain then unidentified.

In our work two special cases of the general MSSV process are of particular interest: the one with a switching intercept and the other — with a regime-dependent autoregression parameter. A concise discussion of both follows.⁶

2.1. MSSV model with a switching intercept, MSSV(μ)

In this case Equation 2 collapses to:

$$\ln h_t = \mu_{S_t} + \varphi \ln h_{t-1} + \sigma \eta_t = \begin{cases} \mu_1 + \varphi \ln h_{t-1} + \sigma \eta_t & \Leftrightarrow S_t = 1 \\ \mu_2 + \varphi \ln h_{t-1} + \sigma \eta_t & \Leftrightarrow S_t = 2 \end{cases}. \quad (3)$$

For the sake of identifiability of the model we reparametrize the switching parameter as (see So et al., 1998):

$$\mu_{S_t} = \gamma_1 + \gamma_2 I(S_t = 2),$$

where $\gamma_1 \in R$, $\gamma_2 < 0$ and $I(\cdot)$ denotes the indicator function which takes one if the condition in the parentheses is satisfied and zero otherwise. Such a representation of the switching intercept results in inequality $\mu_1 > \mu_2$. It may be shown that the latter is equivalent to predetermining states '1' and '2' as ones of high and low mean log-volatility level, respectively, that is:

$$\mu_1 > \mu_2 \Leftrightarrow E(\ln h_t | S_t = 1) > E(\ln h_t | S_t = 2).$$

For the model in question we shall also assume covariance stationarity of the log-volatility process following Equation 3, for which it is necessary and sufficient⁷ to guarantee that $|\varphi| < 1$.

2.2. MSSV model with a switching autoregression parameter, MSSV(φ)

In our study we note that discrete shifts in the mean volatility level may result from not only a switching intercept, but — alternatively — a regime-changing autoregression parameter. Hence, we consider a MSSV model with Equation 2 assuming the form:

$$\ln h_t = \mu + \varphi_{S_t} \ln h_{t-1} + \sigma \eta_t, \quad (4)$$

⁶ For a comprehensive work on non-switching SV models we refer to Pajor (2003).

⁷ General results on second-order and strict stationarity of switching vector autoregression processes may be found in Francq and Zakoian (2001).

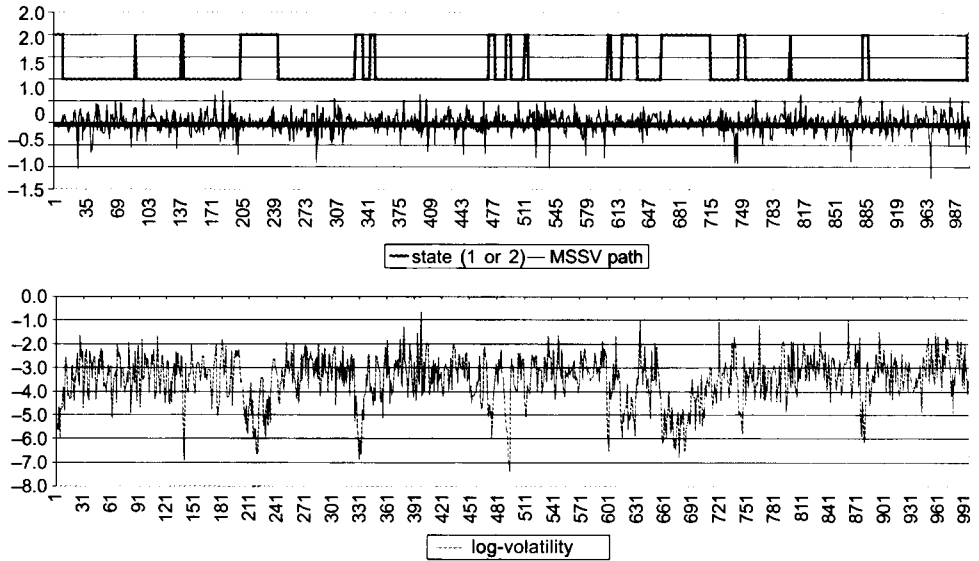


Fig. 1. Simulated path of a MSSV(φ) process ($\mu = -2.5$; $\varphi_1 = 0.2$; $\varphi_2 = 0.5$; $\sigma^2 = 0.6132$; $p_{11} = 0.98$; $p_{22} = 0.95$) and the corresponding log-volatility and regime-switching processes

For the identifiability reasons we shall impose the inequality $\varphi_1 < \varphi_2$. Once formulas for conditional expectations $E(\ln h_t | S_t = i)$ ($i = 1, 2$) are available (see the following subsection), it is easily shown that:

$$\mu < 0 \Rightarrow [\varphi_1 < \varphi_2 \Leftrightarrow E(\ln h_t | S_t = 1) > E(\ln h_t | S_t = 2)]$$

and

$$\mu > 0 \Rightarrow [\varphi_1 < \varphi_2 \Leftrightarrow E(\ln h_t | S_t = 1) < E(\ln h_t | S_t = 2)],$$

which indicates different mean volatility levels in each of the regimes.

Further, we assume covariance stationarity of the log-volatility process following Equation 4. The relevant (necessary and sufficient) condition is given by the set of inequalities:⁸

$$\begin{cases} R_1 < 1 \\ R_2 < 2, \end{cases}$$

where:

$$R_1 = p_{11}\varphi_1^2 + p_{22}\varphi_2^2 + (1 - p_{11} - p_{22})\varphi_1^2\varphi_2^2,$$

$$R_2 = p_{11}\varphi_1^2 + p_{22}\varphi_2^2.$$

⁸ The condition is also valid for the general case, in which all the three volatility parameters are allowed regime shifts (see Francq and Zakoian, 2001). One should note that the condition is somewhat contrary to an initial 'intuition' according to which it should be 'enough' to assume that $|\varphi_1| < 1$ and $|\varphi_2| < 1$. The latter constitutes neither a necessary nor a sufficient condition for second-order stationarity of a two-state switching first-order autoregression (ibid.).

Figure 1 depicts a simulated path of a certain MSSV(φ) process and the corresponding regime-switching and (stationary) log-volatility processes. The latter displays evident shifts in the mean level (according to the switching mechanism), which manifest themselves in the form of volatility clustering.

2.3. Selected regime characteristics

While allowing different states of the economy, it is natural to characterize the regimes in some systematic way. In our work we do so by calculating selected regime-specific characteristics both of the log-volatility process and the switching mechanism as well, including:

— state-conditional mean log-volatility level:⁹

$$E_1 \equiv E(\ln h_t | S_t = 1) = \frac{\mu_1(1 - \varphi_2 p_{22}) + \mu_2 \varphi_1(1 - p_{11})}{1 - \varphi_1 \varphi_2 - \varphi_2 p_{22}(1 - \varphi_1) - \varphi_1 p_{11}(1 - \varphi_2)},$$

$$E_2 \equiv E(\ln h_t | S_t = 2) = \frac{\mu_2(1 - \varphi_1 p_{11}) + \mu_1 \varphi_2(1 - p_{22})}{1 - \varphi_1 \varphi_2 - \varphi_2 p_{22}(1 - \varphi_1) - \varphi_1 p_{11}(1 - \varphi_2)};$$

— state-conditional variance of the log-volatility process:

$$V_i \equiv \text{Var}(\ln h_t | S_t = i) = E(\ln^2 h_t | S_t = i) - E_i^2 \quad \text{for } i = 1, 2,$$

where:

$$E(\ln^2 h_t | S_t = 1) = \frac{d_1(1 - \varphi_2^2 p_{22}) + d_2 \varphi_1^2(1 - p_{11})}{1 - \varphi_1^2 p_{11} - \varphi_2^2 p_{22} + \varphi_1^2 \varphi_2^2(-1 + p_{11} + p_{22})},$$

$$E(\ln^2 h_t | S_t = 2) = \frac{d_2(1 - \varphi_1^2 p_{11}) + d_1 \varphi_2^2(1 - p_{22})}{1 - \varphi_1^2 p_{11} - \varphi_2^2 p_{22} + \varphi_1^2 \varphi_2^2(-1 + p_{11} + p_{22})},$$

and

$$d_i = \mu_i^2 + 2\mu_i \varphi_i E(\ln h_{t-1} | S_t = i) + \sigma_i^2 \quad \text{for } i = 1, 2,$$

$$E(\ln h_{t-1} | S_t = i) = \sum_{j=1}^2 p_{ji}^* E(\ln h_t | S_t = j)$$

with $p_{ji}^* = \text{Pr}(S_{t-1} = i | S_t = j)$ being the inverse transition probabilities;¹⁰

⁹ First- and second-order state-conditional moments of the log-volatility process have been obtained for the general case (that is the one in which all three parameters are regime-changing) under assumption of covariance stationarity of that process and based on the results of Nielsen and Olesen (2000).

¹⁰ In the case of a two-state Markov chain the inverse transition probabilities defined as $p_{ji}^* = \text{Pr}(S_{t-1} = i | S_t = j)$ are easily shown to equal the ordinary transition probabilities, $p_{ji} = \text{Pr}(S_t = i | S_{t-1} = j)$.

— ergodic probabilities:¹¹

$$\pi_1 = \Pr(S_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}},$$

$$\pi_2 = \Pr(S_t = 2) = 1 - \pi_1;$$

— expected duration¹² of state 'i' (once the system has switched to that state; see Hamilton, 1989):

$$Dur_i = \frac{1}{1 - p_{ii}} \quad \text{for } i = 1, 2.$$

3. BAYESIAN ESTIMATION AND COMPARISON OF THE MSSV MODELS¹³

Estimation of the MSSV models is not trivial. Handling the maximum likelihood procedure is riddled with serious numerical obstacles due to the presence of (as much as) two latent processes underlying the observable process: the conditional volatilities, h_t 's, and the states, S_t 's. In our work we resort to Bayesian methodology, which prevails in the MSSV literature.¹⁴ Although new methods — based on the (auxiliary) particle filters — have been developed of recent,¹⁵ we employ the already 'classical' MCMC procedures: the Gibbs sampler and the Metropolis-Hastings algorithm, to simulate from the joint posterior distribution of all the unknown quantities of the model.

Let $y = (y_1, y_2, \dots, y_T)'$ denote the modelled time series, vector $h = (h_1, h_2, \dots, h_T)'$ be the series of the latent conditional volatilities and vector $S = (S_1, S_2, \dots, S_T)'$ — the unobserved Markov chain. We define the parameter vector as $\theta = (\beta', \sigma^2, p_{11}, p_{22})'$ with σ^2 , p_{11} and p_{22} being the parameters common to both MSSV models, and β comprising the model-specific parameters:

$$\beta = \begin{cases} (\gamma_1, \gamma_2, \varphi)' & \text{for } MSSV(\mu) \\ (\mu, \varphi_1, \varphi_2)' & \text{for } MSSV(\varphi) \end{cases}$$

Further, we employ the data-augmentation technique introduced by Tanner and Wong (1987), within which all the unknown quantities of the model

¹¹ Ergodic probabilities calculated for an ergodic Markov chain tell us approximately for how long (in terms of a part of the analyzed time series) the chain remains in each of its states. In the study we assume that $p_{ii} \in (0,1)$ for $i = 1, 2$, which ensures ergodicity of the switching process.

¹² Expected duration of a certain regime is calculated conditionally upon being in that state.

¹³ In the paper we discuss only the estimation of the switching SV models. For a detailed description of the Bayesian estimation of simple SV constructions we refer to Pajor (2003).

¹⁴ For the relevant references see Section 1.

¹⁵ Ditto.

are treated as random variables subject to estimation and taking values in the common space:

$$\omega' = (\theta', h', S') \in \Omega = \Theta \times H \times Q^T,$$

where $\theta \in \Theta \subset R^6$, $h \in H \subseteq R_+^T$, $S \in Q^T$ and $Q = \{1, 2\}$.

The joint posterior distribution of ω is factorized as:¹⁶

$$p(\theta, h, S | y) \propto p(y | h) p(h | S, \theta) p(S | \theta) p(\theta), \quad (5)$$

which reveals its hierarchical structure. Individual components of (5) are presented in the Appendix. Here, we focus on the prior structure of the parameters, for (almost) all of which mutual independence is assumed:

$$p(\theta) = \begin{cases} p(\beta) p(\sigma^2) p(p_{11}) p(p_{22}) & \text{for MSSV}(\mu), \\ p(\beta | p_{11}, p_{22}) p(\sigma^2) p(p_{11}) p(p_{22}) & \text{for MSSV}(\varphi). \end{cases}$$

Conditioning on the transition probabilities in $p(\beta | p_{11}, p_{22})$ for the model with a switching autoregression parameter stems from imposing prior restrictions of covariance stationarity of the underlying log-volatility process (see Section 2.2).

In the study we choose fairly diffuse priors, letting the posterior results arise mainly from the information contained in the data. More specifically, we have:

1. prior distributions for the parameters common to both MSSV models:¹⁷

$$- p(\sigma^2) = f_{IG}(\sigma^2 | v_1, v_2), v_1 = 1, v_2 = 200;$$

$$- p(p_{ii}) = f_B(p_{ii} | a_i, b_i), a_i = b_i = 1, \text{ for } i = 1, 2;$$

where $f_B(p_{ii} | a_i, b_i)$ denotes the density function of a Beta-distributed random variable, p_{ii} , with the shape and scale parameters equal a_i and b_i , respectively;

2. prior distribution for β :

— for the MSSV(μ) model:

$$p(\beta) = f_N^{(3)}(\beta | \beta_0, A_0^{-1}) I(\gamma_2 < 0) I(|\varphi| < 1), \beta_0 = 0_{(3 \times 1)}, A_0 = 0,01 \cdot I_3,$$

where $f_N^{(k)}(\beta | \beta_0, A_0^{-1})$ denotes the density function of a normally distributed k -variate random variable, β , with the mean vector and the precision matrix equal β_0 and A_0 , respectively;

¹⁶ The analysis is conducted conditionally on $h_0 = 1$, dependence on which is omitted in the notation.

¹⁷ We parametrize the density of the inverse gamma distribution as:

$$p(\sigma^2) = f_{IG}(\sigma^2 | v_1, v_2) = \frac{v_2^{-v_1}}{\Gamma(v_1)} \frac{1}{(\sigma^2)^{v_1+1}} \exp\left(-\frac{1}{v_2 \sigma^2}\right).$$

— for the MSSV(φ) model:

$$p(\beta | p_{11}, p_{22}) = f_N^{(3)}(\beta | \beta_0, A_0^{-1})I(R_1 < 0)I(R_2 < 0), \beta_0 = 0_{(3 \times 1)}, A_0 = 0,01 \cdot I_3.$$

As regards prior distributions for the parameters of the BSV model, we follow the structure employed in Pajor (2003), namely:

$$\begin{aligned} &— p(\sigma^2) = f_{IG}(\sigma^2 | v_1, v_2), v_1 = 1, v_2 = 200; \\ &— p(\mu, \varphi) = f_N^{(2)}(\mu, \varphi | 0_{(2 \times 1)}, A_0^{-1})I(|\varphi| < 1), A_0 = 0,01 \cdot I_3. \end{aligned}$$

The prior structure presented above provides very convenient (in terms of the sampling method) conditional posterior distributions of the model parameters.¹⁸ The latter are employed to construct a hybrid chain within the MCMC procedure, through which a N -sized sample from the joint posterior distribution is obtained, $\{\omega^{(q)}\}_{q=M+1}^{M+N}$, where q denotes the number of the cycle of the sampling algorithm, of which the first M cycles are discarded, and $\omega^{(q)}$ signifies the outcome on ω from the q -th step. Once the algorithm is complete, it is straightforward to obtain also a sample of any measurable function of ω , such as regime characteristics, in particular.

In order to allow Bayesian model comparison, the marginal likelihood for each of the estimated models needs to be evaluated. In our work we resort to the procedure introduced by Newton and Raftery (1994), in which the quantity of interest is estimated as:

$$\hat{p}(y | M_i) = \left[\frac{1}{N} \sum_{q=M+1}^{M+N} \frac{1}{p(y | \omega_{(i)}^{(q)}, M_i)} \right]^{-1},$$

where $\hat{p}(y | M_i)$ is the estimator of the marginal likelihood in the i -th model, M_i . Despite its lamentable numerical instability, the method proved satisfactory in our applications. Finally, to compare the models pair-wise use is made of Bayes factors, B_{ij} , defined as:

$$B_{ij} = \frac{p(y | M_i)}{p(y | M_j)}.$$

¹⁸ Full details on the posterior structure of all the estimated quantities are found in the Appendix.

4. EMPIRICAL STUDY

4.1. Data description

The methodology presented above is illustrated with an empirical study in which data from the Polish financial market is analyzed. More specifically, we consider a series of daily quotations of the 1-month Warsaw Interbank Offered Rate (WIBOR1M) interest rates over the period from April 17, 2000 to April 7, 2008 (which makes the total of 2002 observations). The series is plotted in Figure 2.

WIBOR1M interest rates (2000.04.17–2008.04.07)

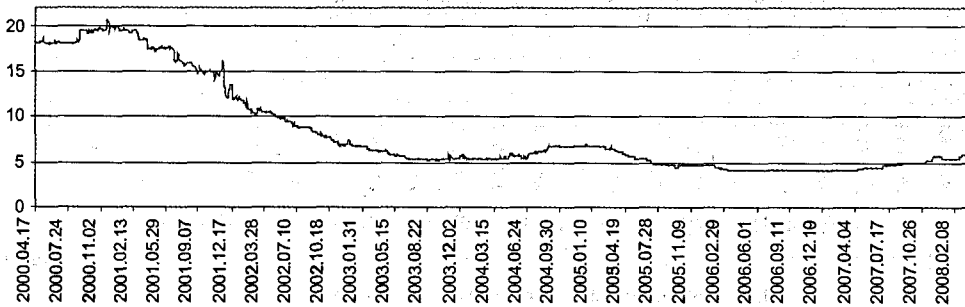


Fig. 2. The series of WIBOR1M interest rates, w_t ,

We calculate the daily log-returns, r_t 's, on the WIBOR1M interest rates, defined as:

$$r_t = 100 \ln(w_t / w_{t-1}),$$

where w_t denotes the price of the instrument at time t . The series of r_t 's is presented in Figure 3.

Log-returns on WILBOR1M (2000.04.18–2008.04.07)

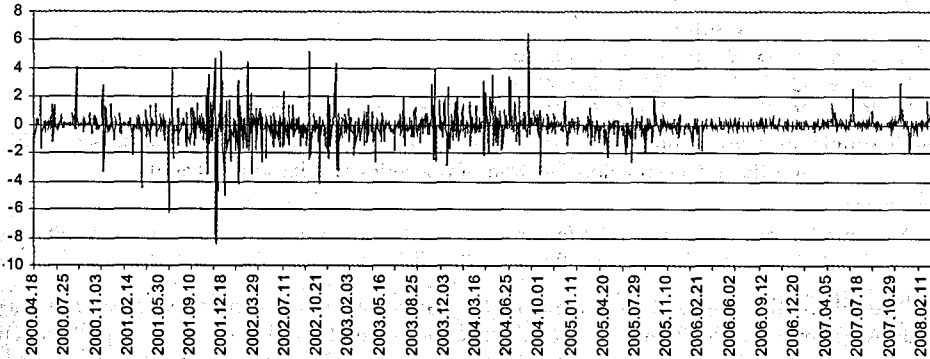


Fig. 3. Daily log-returns, r_t , on the WIBOR1M (April 18, 2000–April 7, 2008)

Further, we apply a simple linear filter — a first-order autoregressive model — to the data as to account for possibly non-zero conditional mean of the data generating process.¹⁹ Henceforth, the analysis is conducted for the resulting series of AR(1)-residuals²⁰ (see Fig. 4), denoted as y_t with $t = 1, 2, \dots, T = 2000$. The latter display features commonly found in financial data, including volatility clustering, high value of the empirical kurtosis coefficient (see Tab. 1), no

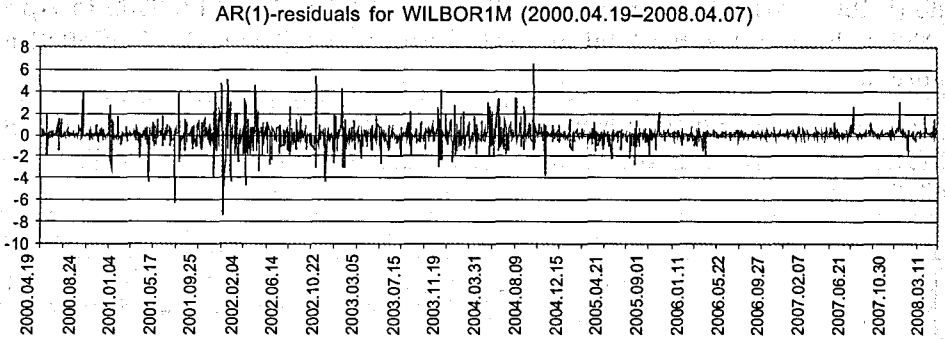


Fig. 4. The AR(1)-residuals for the daily log-returns on the WIBOR1M

Table 1

Descriptive statistics for AR(1)-residuals for WIBOR1M

Min	Max	Mean	Stand. deviation	Asymmetry	Kurtosis	ARCH(2) effect
-7.2629	6.5844	0.0000	0.8394	-0.4339	18.5818	$TR^2 = 177.1391$ ($p\text{-value} = 0.0000$)

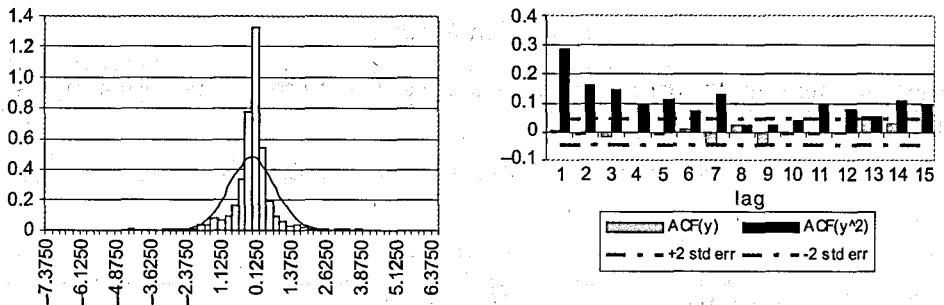


Fig. 5. Empirical distribution of the series $\{y_t, t = 1, 2, \dots, T\}$ with fitted normal density (left) and the autocorrelation function of the series and its square (right)

¹⁹ Estimation of the AR(1) model for the series of the log-returns, r_t 's, yields the results:

$$r_t = -0.0464 + 0.1537r_{t-1} + \hat{u}_t$$

(0.0188) (0.0221)

²⁰ An alternative approach is to simulate the parameters of the conditional mean modelled with an AR(1) process from their conditional posterior distributions. However, we expect that autocorrelations in the log-returns have little impact on the volatility and, hence, adopt the method used by So et al., 1998.

significant autocorrelations in the original series, yet strong auto-dependencies in the squared series (see Fig. 5). Additionally, left asymmetry in the empirical distribution of the residuals is found (see Tab. 1).

4.2. Results for the Basic SV model

For the estimation of the BSV model we employed the Gibbs sampler combined with the Metropolis-Hastings step for sampling the latent conditional volatilities, h_t 's, as done in Pajor (2003). The first $M = 500,000$ burnt-in iterations are discarded and the subsequent $N = 1,500,000$ iterations are regarded as a simulated sample from the joint posterior density.

Table 2 contains posterior means and standard deviations of the model parameters. One notes the posterior mean of the autoregression parameter, φ , being fairly close to unit. It is a common finding in the SV literature (see Pajor, 2003, among many), indicating evident persistence in the conditional volatility process. Despite prior independency between the parameters we observe strong posterior correlations (see Tab. 3). The latter may arise as a result of 'stabilization' of the unconditional characteristics of the volatility process, such as mean and variance.

Table 2

Posterior means and standard deviations (in parentheses) of the BSV parameters

μ	φ	σ^2
-0.3384 (0.0508)	0.8269 (0.0215)	1.1053 (0.1285)

Table 3

Posterior correlation matrix of the BSV parameters

M_1	μ	φ	σ^2
μ	1	0.8753	-0.6590
φ		1	-0.7185
σ^2			1

In Figure 6 the marginal prior and posterior distributions of the BSV parameters along with the plots of their ergodic means (against the number of cycles) are depicted. The results of posterior densities being of regular shapes and fast convergence of the ergodic means convergence to their posterior counterparts remain consistent with Pajor (2003).

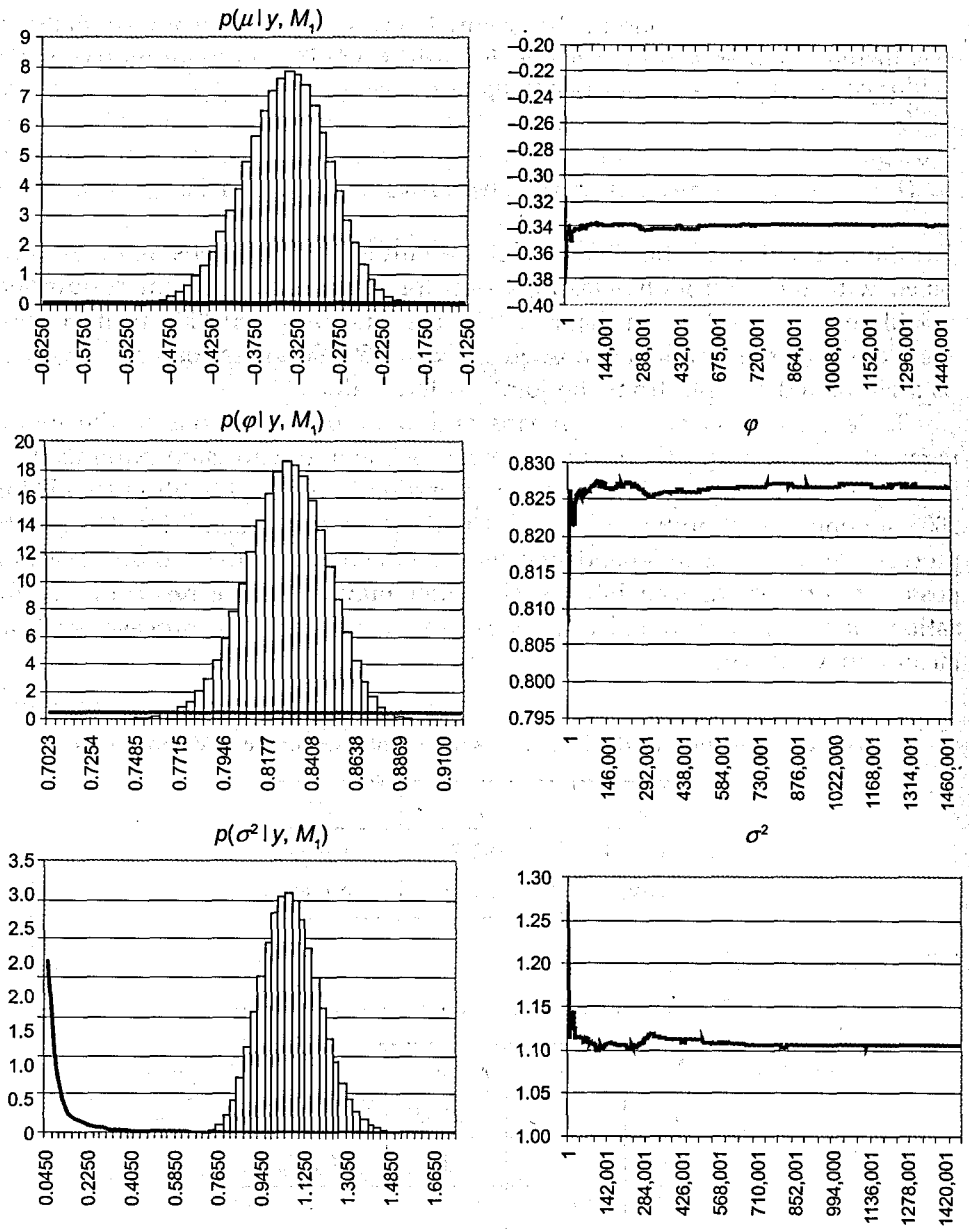


Fig. 6. Left column: marginal prior (solid line) and posterior distributions of the BSV parameters; right column: ergodic means of the parameters against the number of cycles

4.3. Results for the MSSV (μ) model

To estimate the model we employ the sampling algorithm presented in the Appendix.²¹ The first $M = 2,000,000$ burn-in iterations are discarded and the subsequent $N = 1,500,000$ iterations constitute a simulated sample from the joint posterior density.

As it can be gathered from the posterior means of the transition probabilities located very close to unit (see Tab. 4), the switching mechanism manifests strong persistence. Once a certain state is achieved, little is the probability of a switch to the alternative regime. Furthermore, one notes significantly negative posterior mean of γ_2 , which provides compelling evidence of discrete shifts in the value of the intercept. As compared with the results for the BSV model, the mean posterior of the elasticity of volatility is markedly lower. It is the most common finding cited in the MSSV literature, where it is argued that structural shifts unaccounted for by standard SV models may imply spuriously high persistence in the volatility process. However, we would not jump to such conclusions, unless the true autocorrelation functions of the log-volatility process in the BSV and MSSV model are surveyed.²² One may presume, that the very same 'spuriously' high volatility persistence implied by the non-switching SV specification may be captured by the switching counterpart, yet in a different manner (resulting, for instance, in the close-to-unit mean posterior probabilities p_{ii} , $i = 1, 2$). The issue merits further research.

Table 4

Posterior means and standard deviations of the parameters of model M_2

p_{11}	p_{22}	γ_1	γ_2	φ	σ^2
0.9960	0.9964	-0.2753	-0.7292	0.6658	1.3594
(0.0027)	(0.0026)	(0.0647)	(0.1051)	(0.0360)	(0.1437)

According to the posterior correlation matrix of the parameters (see Tab. 5), prior assumption of their mutual independence seems to be overruled by the data. In our belief, the non-zero posterior correlation coefficients may arise from 'stabilization' of the regime characteristics as well as the unconditional moments of the log-volatility process.

²¹ We note that the minimum acceptance rate while sampling h_t 's via the M-H algorithm, amounted to approximately 60%, which is found much satisfying.

²² For the purpose of comparison of volatility persistence implied by the BSV and both MSSV models, we averaged posterior empirical autocorrelation function (ACF) coefficients (lags: 1 to 15) for the sampled series of $\ln h_t$'s. The results appear not to reject the individual hypotheses of equal mean ACF coefficients across different models, therefore advocating the conjecture to follow in the main text.

Posterior correlation matrix of the parameters of model M_2

M_2	p_{11}	p_{22}	γ_1	γ_2	φ	σ^2
p_{11}	1	0.3966	-0.0062	0.1986	0.1292	0.0035
p_{22}		1	0.1524	0.2232	0.1523	0.0078
γ_1			1	0.1262	0.5273	-0.3921
γ_2				1	0.8012	-0.4871
φ					1	-0.6647
σ^2						1

Marginal posterior densities of the transition probabilities p_{ii} concentrate tightly on the left of unit (see Fig. 7), which indicates that the analyzed dataset is very informative with regard to the switching mechanism. Posterior distri-

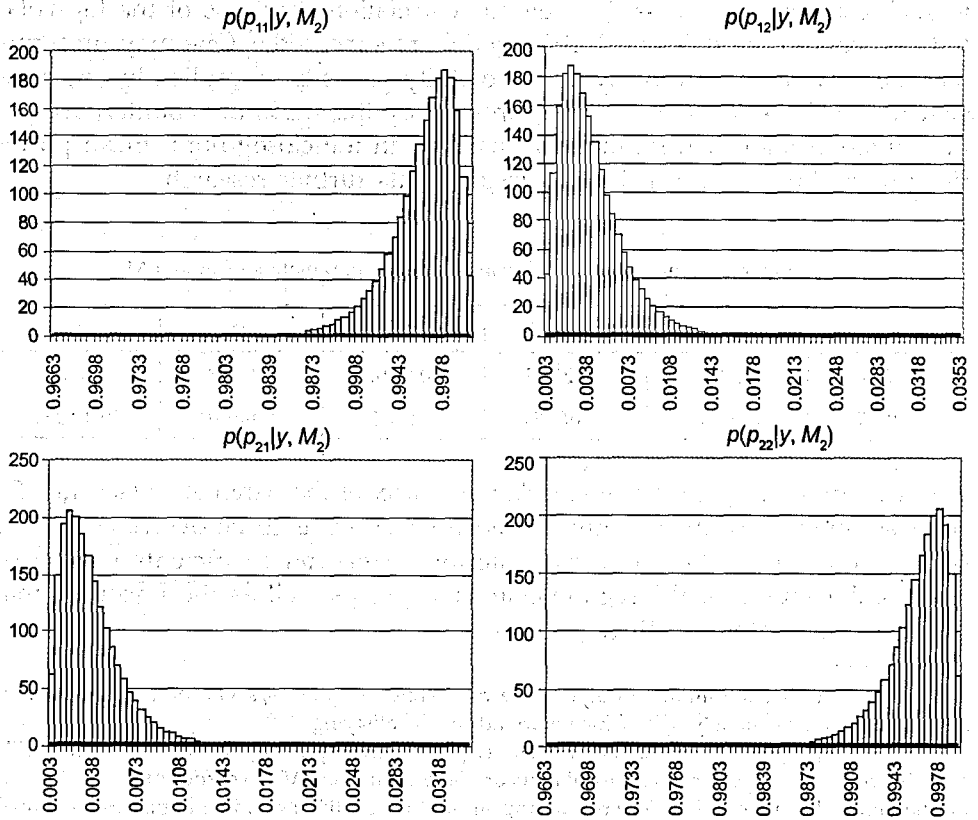


Fig. 7. Marginal prior (solid line) and posterior distributions of the transition probabilities in model M_2

butions of the remaining parameters are clearly unimodal and cluster (with slight asymmetries) around their means (see Fig. 8). Prior covariance stationarity of the log-volatility process is not rejected by the data, as the posterior density of the autoregression parameter, ϕ , clusters away on the left of unit.

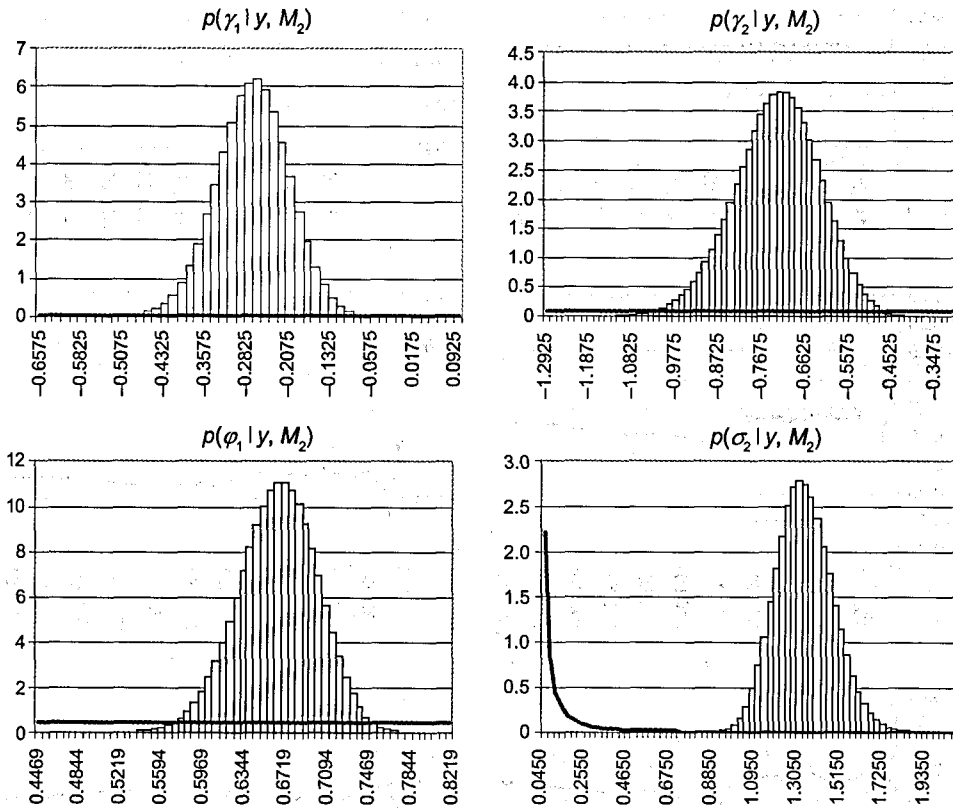


Fig. 8. Marginal prior (solid line) and posterior distributions of the log-volatility parameters in model M_2

A somewhat unstable behaviour of the ergodic means of the parameters (except for the transition probabilities) may raise concerns as regards convergence of the MCMC procedure (see Fig. 9). However, one should note rather negligible magnitude of the visible fluctuations.

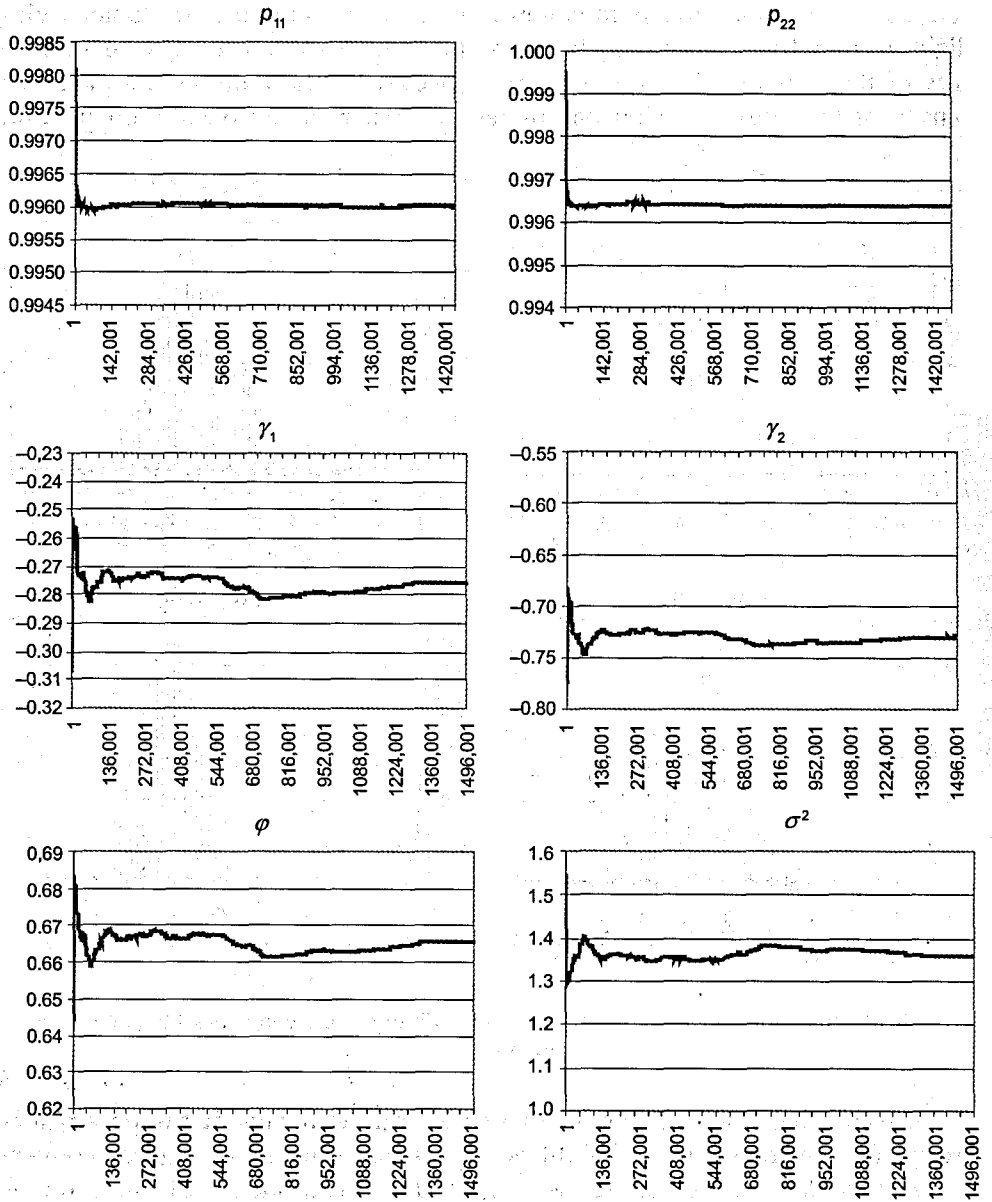


Fig. 9. Ergodic means of the parameters of model M_2 against the number of cycles

4.4. Results for the MSSV(φ) model

As in the previous case, the quantities of interest are sampled within the MCMC procedure presented in the Appendix.²³ The first $M = 2,000,000$ burn-in iterations are discarded and the subsequent $N = 1,500,000$ iterations constitute a simulated sample from the joint posterior density.

As far as the switching mechanism is concerned, similar (to the previous model) results are obtained. Posterior means of the probabilities p_{ii} are very

Table 6

Posterior means and standard deviations of the parameters of model M_3

p_{11}	p_{22}	μ	φ_1	φ_2	σ^2
0.9939	0.9961	-0.4511	0.5465	0.8201	1.2485
(0.0046)	(0.0026)	(0.0642)	(0.0869)	(0.0233)	(0.1375)

close to unit, implying high persistence in the latent Markov chain (see Tab. 6). Moreover, the posterior means of the switching parameter differ substantially across the two regimes. It follows that switches between two genuinely dis-

Table 7

Posterior correlation matrix of the parameters of model M_3

M_3	p_{11}	p_{22}	μ	φ_1	φ_2	σ^2
p_{11}	1	0.4241	0.0631	0.3608	0.1015	0.0225
p_{22}		1	0.0970	0.0765	0.0071	-0.0020
μ			1	0.4967	0.8355	-0.6405
φ_1				1	0.5678	-0.3328
φ_2					1	-0.6360
σ^2						1

²³ The minimum acceptance rate while sampling h_i 's via the M-H algorithm, amounted to approximately 60%.

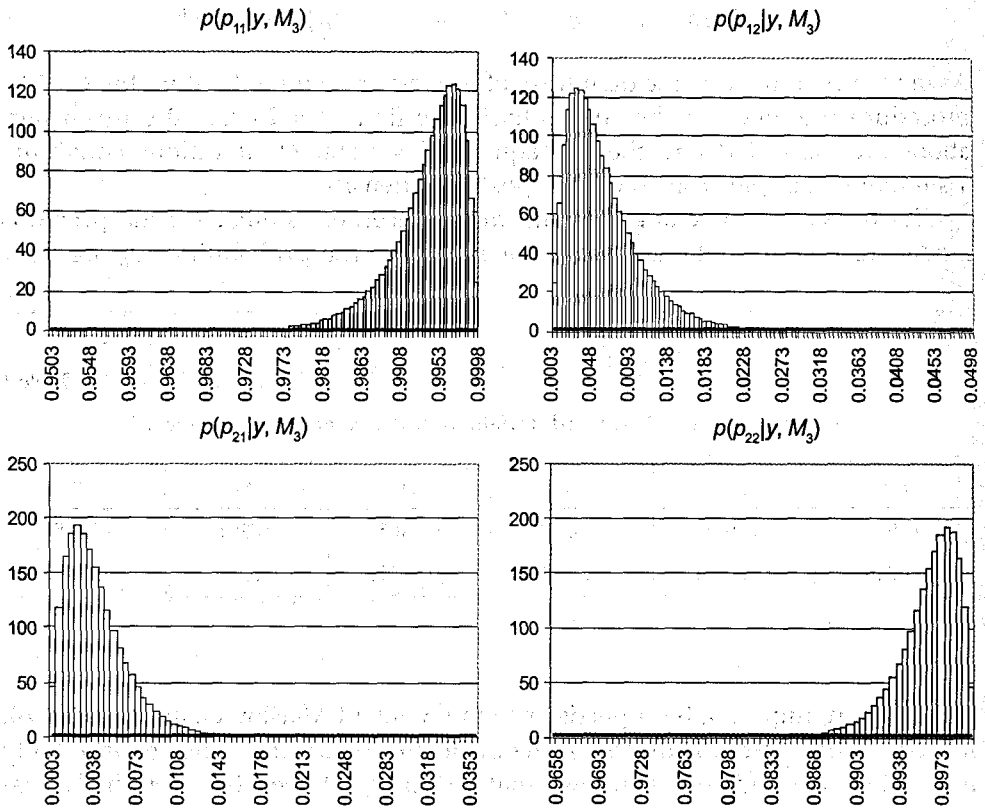


Fig. 10. Marginal prior (blue solid line) and posterior distributions of the transition probabilities in model M_3

tinct states of the economy do occur in the modelled time series. Again, the posterior correlations between the parameters appear to reject their prior independence, a reason for which is believed to be the same as in the model with a regime-changing intercept.

Marginal posterior densities of the transition probabilities resemble much those obtained for the MSSV(μ) model. Apart from a strong left asymmetry of parameter φ_1 , no other irregularities are found in the posterior distributions of the parameters (see Fig. 11).

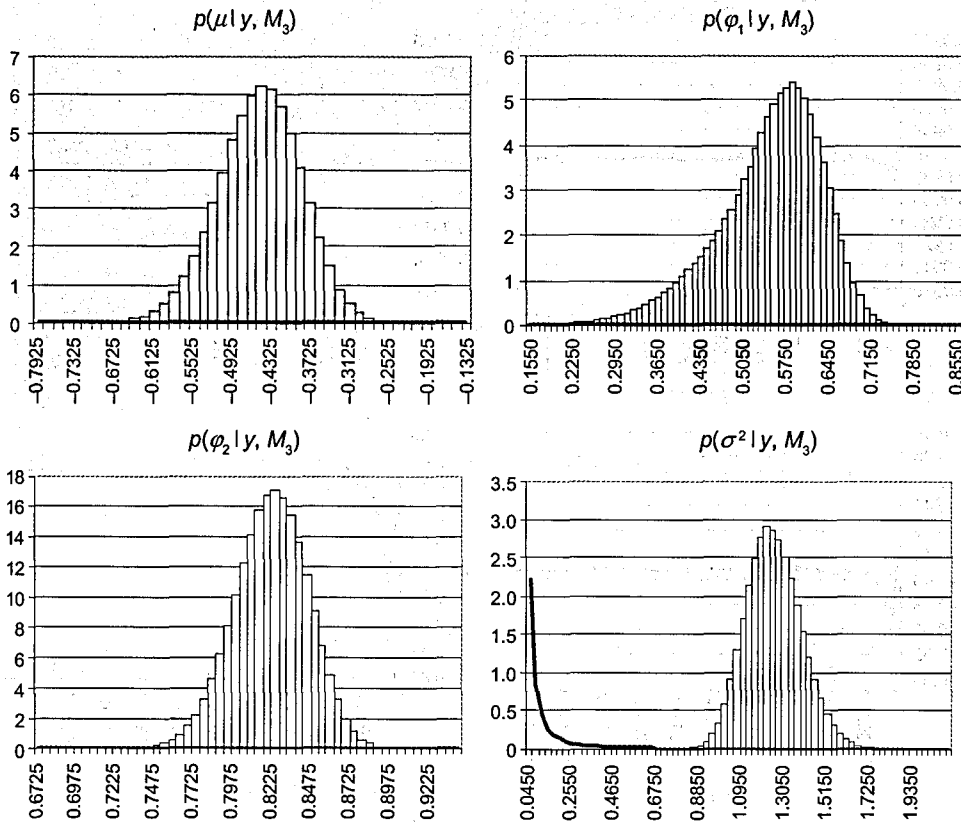


Fig. 11. Marginal prior (solid line) and posterior distributions of the log-volatility parameters in model M_3

The behaviour of the ergodic means seems to raise no concerns with regard to the convergence of the MCMC algorithm (see Fig. 12).

To analyze the validity of the prior constraints for second-order stationarity of the log-volatility process is a more demanding task than in the previous cases. Therefore, we present Figure 13, plotting the values of R_1 and R_2 , which are required to satisfy the inequalities: $R_1 < 1$ and $R_2 < 2$ (see Section 2.2). We note that only the dark area in the figure represents the set of pairs (R_1, R_2) that guarantee stationarity of the log-volatility process.²⁴ Within the region two-dimensional contours of the posterior density of random vector (R_1, R_2) are plotted. Despite the location of the latter close to the stationarity border, the data appears rather not to reject the prior stationarity restrictions.

²⁴ The stationarity region has been obtained by simulation and therefore displays slight inaccuracies.

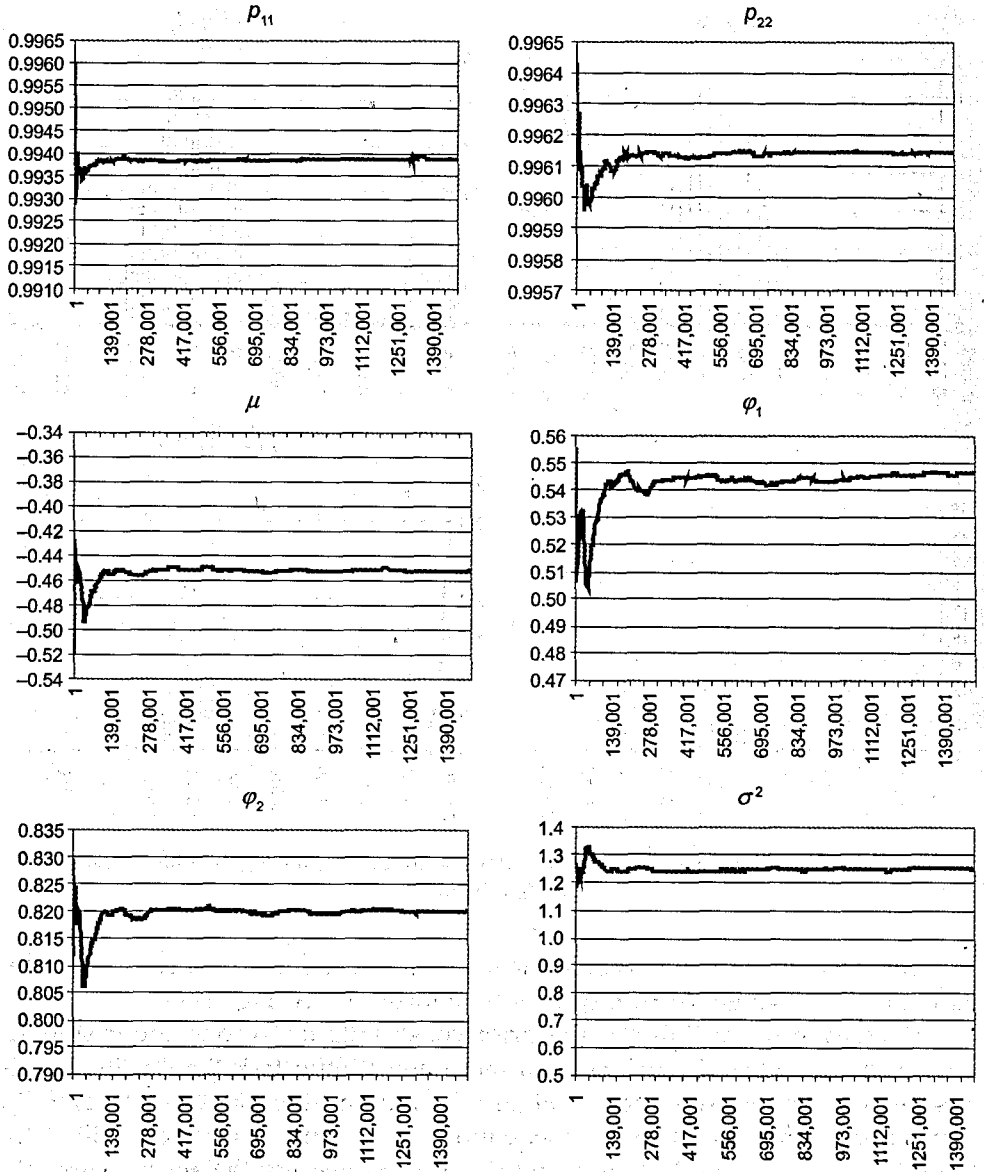


Fig. 12. Ergodic means of the parameters of model M_3 against the number of cycles

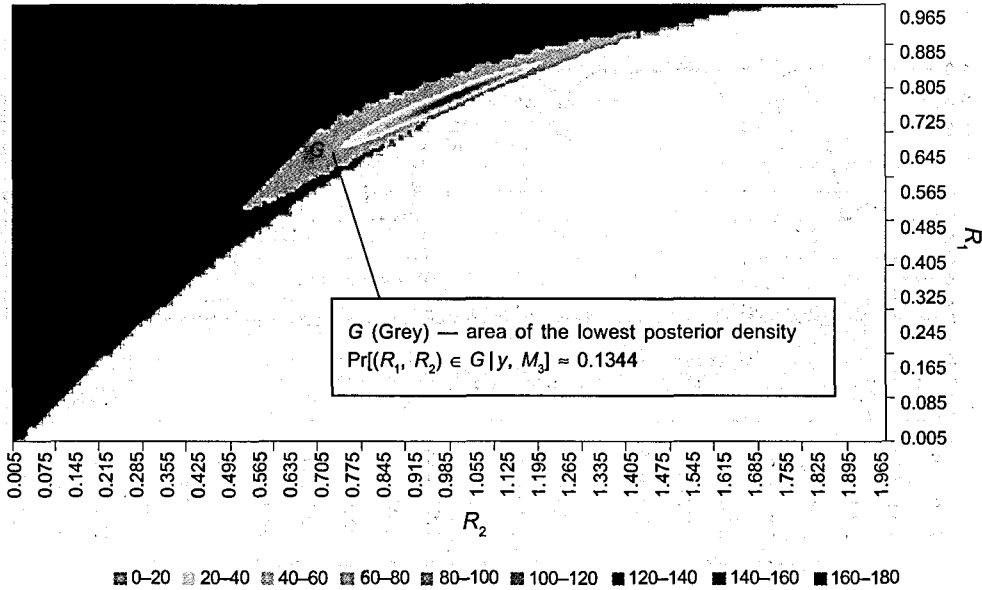


Fig. 13. Simulated region of possible values of restrictions ($R_1 < 1, R_2 < 2$) for covariance stationarity of a MSSV(ϕ) process (dark area) along with contours of their bivariate posterior distribution obtained for the data

4.5. Regime characteristics

Both regime-switching SV specifications imply existence of two distinguishable states of the economy. It is evident even more in Figure 14, depicting averaged posterior probabilities²⁵ $\Pr(S_t = 1 | y)$ in each of the two models along with the modelled time series and the averaged posterior log-volatilities, $\ln h_t$'s, extracted from model M_2 .²⁶ Unit-close values of the 1-state mean probabilities clearly correspond with the period of relatively higher volatility of the daily WIBOR1M interest rates (from about April, 2001 to September, 2004). Most of the remaining part of the sample period is definitely labelled as state '2'. There is a rather short sub-period, however, lasting from March, 2005 to September, 2005, that cannot be ascribed to any of the regimes unambiguously. It may be

²⁵ Mean posterior probabilities of state '1' have been obtained as:

$$\Pr(S_t = 1 | y) = \frac{1}{N} \sum_{q=M+1}^{N+M} I(S_t^{(q)} = 1), \quad t = 1, 2, \dots, T.$$

²⁶ The series of the averaged posterior $\ln h_t$'s only from model M_2 is presented, as it coincides quite much with the ones obtained from other specifications, i.e. M_1 and M_3 .

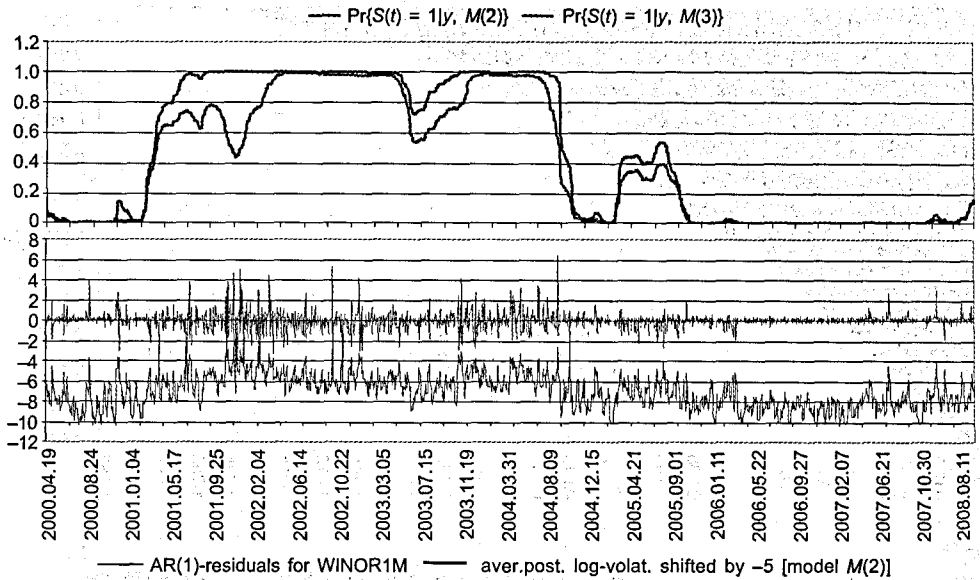


Fig. 14. Mean posterior probabilities of state '1' in models M_2 and M_3 (upper plot) along with the modelled time series (lower plot) and averaged posterior $\ln h_t$'s model M_2

case that yet another state (i.e. the one representing a medium volatility level) should be introduced to the model, yielding a three-state MSSV specification.

Posterior means of the regime characteristics (see Tab. 8) indicate that the models differentiate the two regimes in terms of either only the mean log-volatility level (model M_2) or, additionally, in terms of the state-dependent variances of the log-volatility process (model M_3), with the low-volatility state '2' featuring relatively increased 'variability of volatility'. One should note that what characterizes the expected durations of each of the states is considerable dispersion (in terms of the standard deviation) in their posterior densities featuring very long and heavy right tails (see Fig. 16). On average, the expected time of the Markov chain remaining in a particular state (once it has been achieved) differs from its posterior mean by about 805 to 5251 weekdays (see Tab. 8). As regards the ergodic probabilities, it is noticed that their posterior densities, though of a regular shape, are fairly diffused over the unit interval, therefore precluding precise inference on approximately how long²⁷ the chain remains in a particular state.²⁸ On the other hand, posterior distributions of the remaining regime characteristics (i.e. state-dependent log-volatility means and variances) evidently cluster around their posterior means, although slight asymmetries in their profiles may be observed (see Fig. 17 and 18).

²⁷ In terms of a part of the entire period over which the data is analyzed.

²⁸ We draw attention to the fact, that such an interpretation of the ergodic probabilities of a Markov chain is valid once the chain has converged to its stationary (ergodic, invariant) distribution.

Table 8

Posterior means and standard deviations (in parentheses) of selected regime characteristics in model M_2 and M_3

Model	Regime characteristics							
	π_1	π_2	Dur_1	Dur_2	E_1	E_2	V_1	V_2
M_2 [MSSV(μ)]	0.4736 (0.1873)	0.5264 (0.1873)	467.60 (1149.07)	555.61 (5251.38)	-0.8396 (0.1639)	-2.9871 (0.1615)	2.4658 (0.1967)	2.4643 (0.1964)
M_3 [MSSV(ϕ)]	0.4104 (0.1799)	0.5896 (0.1799)	337.52 (805.24)	484.60 (979.22)	-1.0254 (0.1600)	-2.4847 (0.1959)	1.8301 (0.2428)	3.8422 (0.3792)

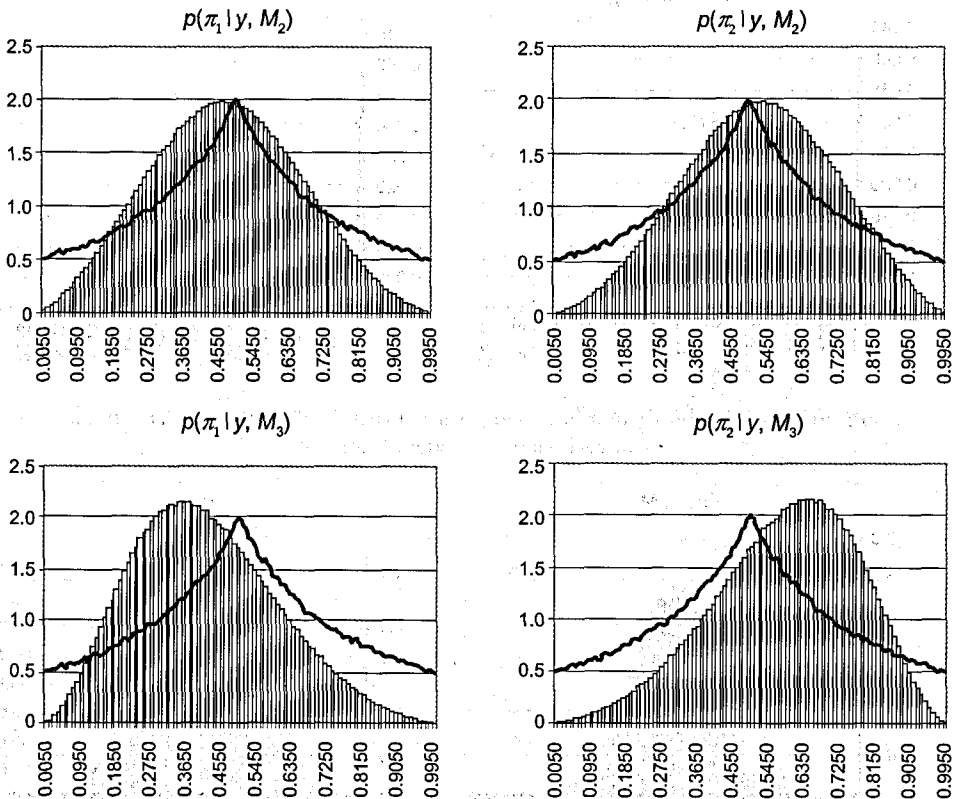


Fig. 15. Marginal prior (solid line) and posterior distributions of ergodic probabilities in model M_2 and M_3

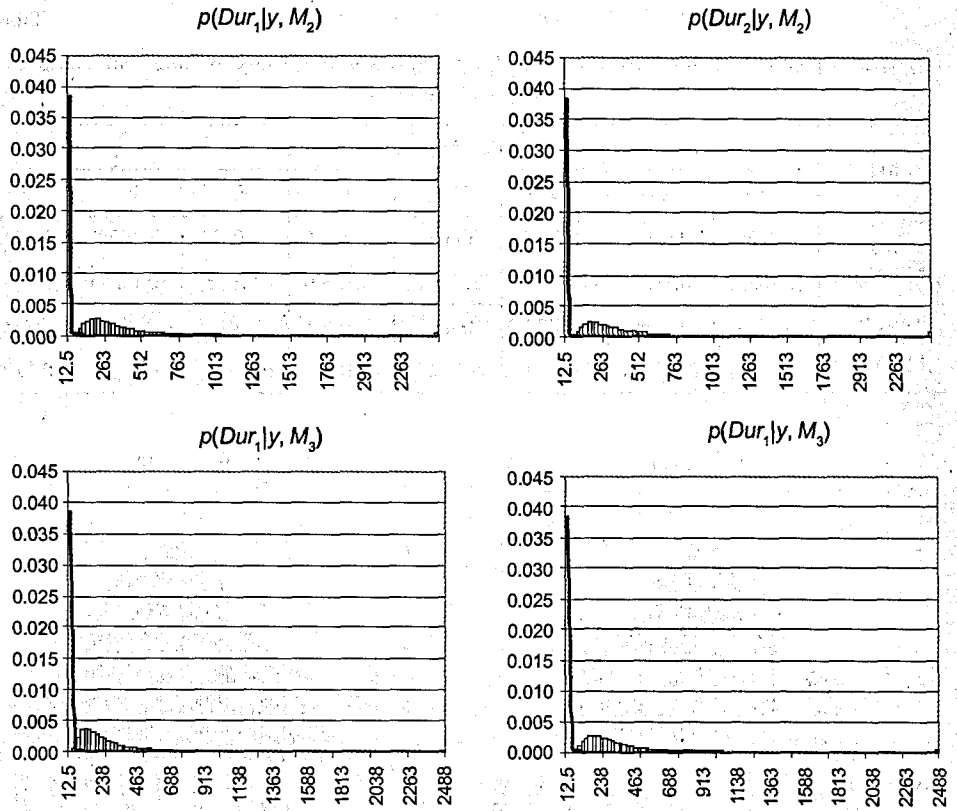


Fig. 16. Marginal prior (solid line) and posterior distributions of expected duration of each state in model M_2 and M_3

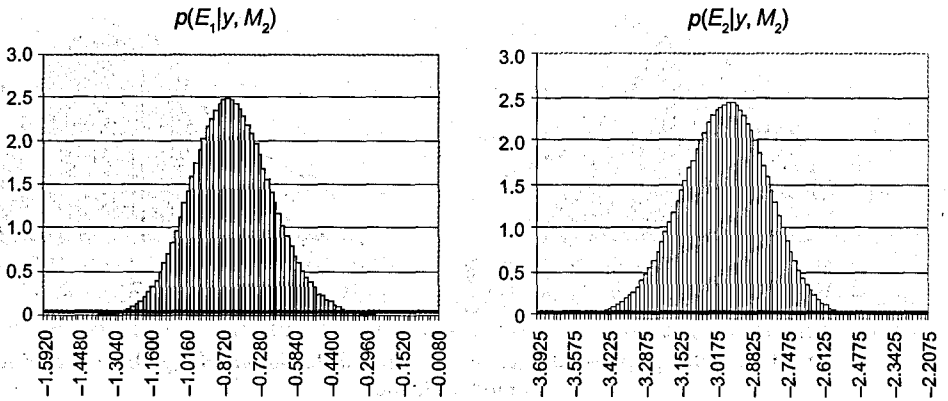


Fig. 17a. Marginal prior (solid line) and posterior distributions of state-conditional log-volatility means in model M_2

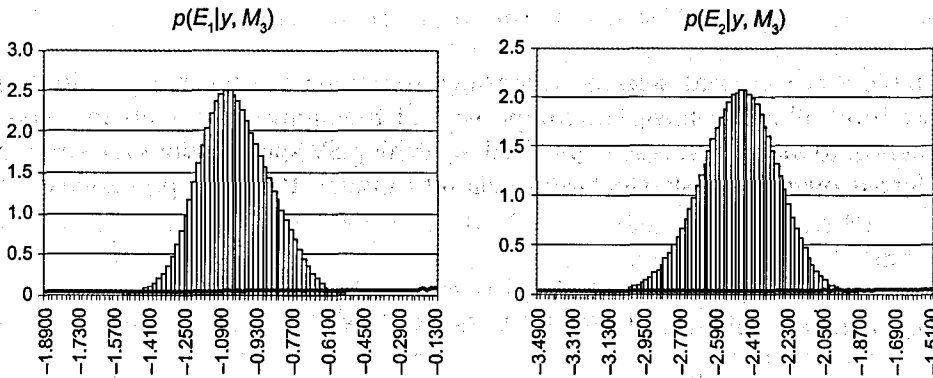


Fig. 17b. Marginal prior (solid line) and posterior distributions of state-conditional log-volatility means in model M_3

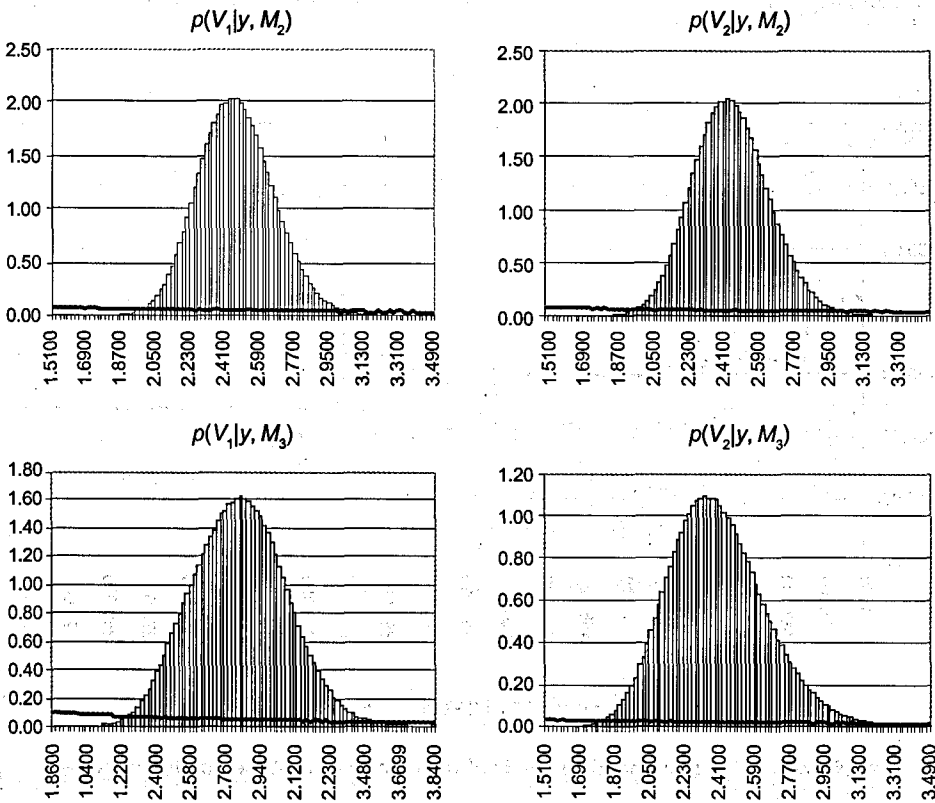


Fig. 18. Marginal prior (solid line) and posterior distributions of state-conditional log-volatility variances in model M_2 and M_3

4.6. Bayesian model comparison

In Table 9 we present selected quantities (obtained via the Newton-Raftery procedure) allowing Bayesian comparison of the analyzed models in respect of their fit to the data. It is seen that both switching SV specifications are strongly preferred over the basic stochastic volatility model. Posterior probability²⁹ of

Table 9

Logs of the marginal likelihoods along with the posterior model probabilities and logs of Bayes factors against model M_1

Model	Number of parameters	$\log_{10} \hat{p}(y M_i)$	$\Pr(M_i y)$	$\log_{10} B_{i1}$	Rank
M_1 (BSV)	3	-440.8213	6.215E-16	0	3
M_2 [MSSV(μ)]	6	-425.6148	0.999997	15.2066	1
M_3 [MSSV(φ)]	6	-431.0917	0.000003	9.7296	2

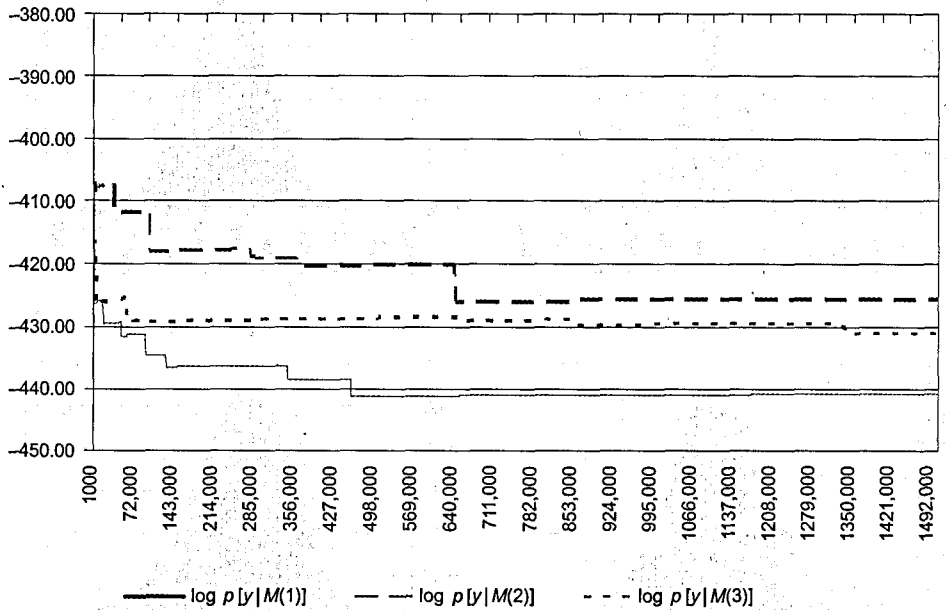


Fig. 19. Logs of the marginal likelihoods of the models against the number of cycles

²⁹ Posterior model probabilities, $\Pr(M_i | y)$, have been obtained under equal prior probabilities of the models, i.e. $\Pr(M_i) = 1/3$.

the model with a switching intercept is as much as about 10^{15} times the posterior chances of the BSV model, and about 10^5 times the chances of the other switching model. These are compelling arguments against the homogeneity (i.e. the lack of structural shifts) of the modelled time series.

Nevertheless, the results may be considered somewhat dubious in view of the notorious instability of the Newton-Raftery algorithm. Therefore, the logs of the marginal likelihood in each of the models and selected Bayes factors are plotted against the number of MCMC iterations (see Fig. 19 and 20). We observe relative stabilization of these quantities only after about 650,000 cycles. More importantly, however, the ranking of the models remains visibly unchanged throughout (see Fig. 19).

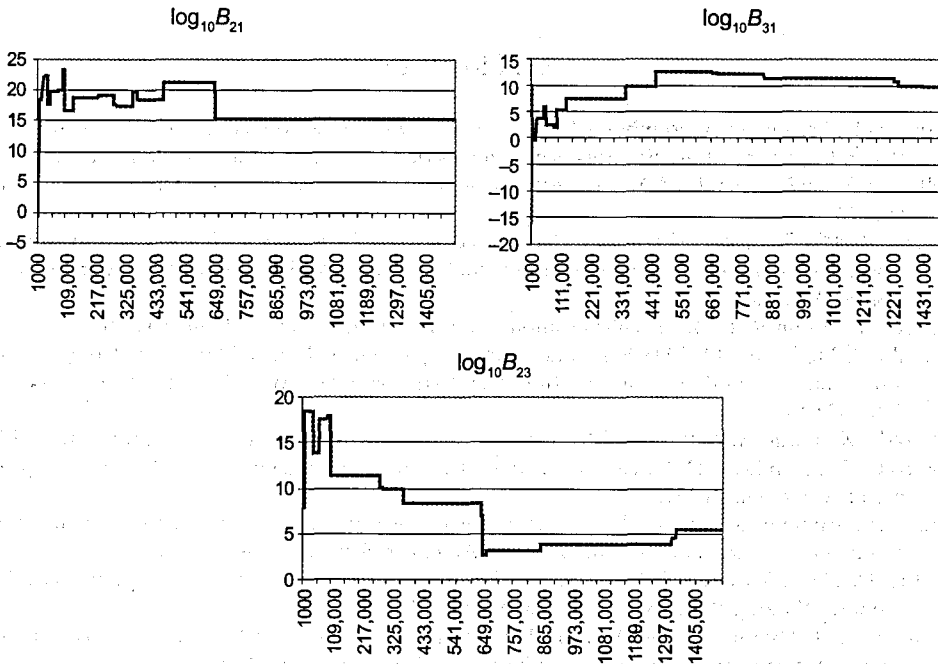


Fig. 20. Logs of Bayes factors against the number of cycles

5. CONCLUSIONS

In the paper two special cases of a general Markov switching SV model are under consideration. One of them allows discrete shifts only in the intercept, whereas the other — in the autoregression parameter of the latent log-volatility process. Both constructions are capable of accounting for sudden changes

in the mean volatility level. Hence, we aim to compare these two specifications in respect of goodness of their fit to the data.

The results of the Bayesian analysis of both switching models as well as a basic SV model provides compelling evidence against homogeneity of the series of the AR(1)-residuals for the daily WIBOR1M interest rates, as evident superiority of the switching models over the BSV construction is observed. Among them the one that features a regime-changing intercept is undoubtedly preferred the most.

The two regimes are distinguishable in terms of either only the mean log-volatility level (in the model with a switching intercept) or, additionally, the state-dependent variances of the log-volatility process (while the autoregression parameter is allowed regime-changing).

REFERENCES

- Bauwens L., Preminger A., Rombouts J. 2006. Regime Switching GARCH Models, Core Discussion Paper, Département des Sciences Économiques de l'Université catholique de Louvain.
- Bollerslev T. 1987. Generalised Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31.
- Casarin R. 2003. Bayesian Inference for Generalised Markov Switching Stochastic Volatility Models, Conference materials at the 4th International Workshop on Objective Bayesian Methodology, CNRS, Aussois.
- Carter C.K., Kohn R. 1994. On Gibbs sampling for state space models, *Biometrika* 81, 3.
- Carvalho C.M., Lopes H.F. 2006. Simulation-based sequential analysis of Markov switching stochastic volatility models, *Computational Statistics & Data Analysis*, doi: 10.1016/j.csda.2006.07.019.
- Diebold F.X., Inoue A. 2001. Long Memory and Regime Switching, *Journal of Econometrics*, 105.
- Franco C., Zakoian J.-M. 2001. Stationarity of multivariate Markov-switching ARMA models, *Journal of Econometrics* 102.
- Frühwirth-Schnatter S. 2001. Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models, *Journal of the American Statistical Association* 96.
- Gärtner D. 2007. Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, Working Paper No. 0719, University of Zurich.
- Granger C.W.J., Hyung N. 1999. Occasional Structural Breaks and Long Memory, Discussion Paper 99-14, Department of Economics, University of California, San Diego.
- Hamilton J.D. 1989. A New approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 2.
- Hwang S., Satchell S. E., Pereira P. L. V. 2003. Stochastic Volatility Models with Markov Regime Switching State Equations, *Journal of Business and Economic Statistics* 16, 2.
- Hwang S., Satchell S. E., Pereira P. L. V. 2004. How Persistent is Volatility? An Answer with Stochastic Volatility Models with Markov Regime Switching State Equations, CEA@Cass Working Paper Series, <http://www.cass.city.ac.uk/cea/index.html>
- Kalimpalli M., Susmel R. 2001. Regime-switching stochastic volatility and short-term interest rates, CEMA Working Papers, <http://ideas.repec.org/p/cem/doctra/197.html>
- Krolzig, H.-M. 1997. Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis, *Lecture Notes in Economics and Mathematical Systems*, New York/Berlin/Heidelberg: Springer.

- Newton M.A., Raftery A.E. 1994. Approximate Bayesian inference by the Weighted Likelihood Bootstrap with Discussion. *Journal of the Royal Statistical Society B* 56, 1.
- Nielsen S., Olesen J. O. 2000. Regime-switching stock returns and mean reversion, Working paper 11-2000, Institut for Nationalokonomi, <http://citeseer.ist.psu.edu>
- Shibata M., Watanabe T. 2005. Bayesian analysis of a Markov switching stochastic volatility model, *Journal of Japan Statistical Society* 35, 2.
- So M. K. P., Lam K., Li W. K. 1998. A stochastic volatility model with Markov switching, *Journal of Business and Economic Statistics* 16, 2.
- Pajor A., 2003. Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych (Stochastic Volatility Processes in Bayesian analysis of financial time series), Cracow University of Economics, Cracow 2003.
- Smith D. R., 2000. Markov-switching and stochastic volatility diffusion models for short-term interest rates, <http://citeseer.ist.psu.edu>
- Tanner M.T., Wong W. 1987. The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association* 82.

APPENDIX

Under the notation established in Section 3, the joint posterior distribution of all the unknown quantities of the MSSV model is decomposed as:

$$p(\theta, h, S | y) \propto p(y | h)p(h | S, \theta)p(S | \theta)p(\theta),$$

Individual components of the above factorization presents themselves as follows:

$$- p(y | h) = \prod_{i=1}^T f_N(y_i | 0, h_i),$$

where: $f_N(y_i | 0, h_i)$ denotes the density function of a normally distributed random variable y_i with mean and variance equal zero and h_i , respectively;

$$- p(h | S, \theta) = \prod_{i=1}^T f_{LN}(h_i | m_i, \sigma^2),$$

where: $f_{LN}(h_i | m_i, \sigma^2)$ denotes the density function of a log-normally distributed random variable h_i with the scale parameter equal σ^2 and the location parameter equal m_i given as:

$$m_i = \begin{cases} \mu_{S_i} + \varphi \ln h_{i-1} & \text{for MSSV } (\mu) \\ \mu + \varphi_{S_i} \ln h_{i-1} & \text{for MSSV } (\varphi) \end{cases}$$

$$- p(S | \theta) \propto p(S_0)p(S | \theta) = p(S_0) \prod_{i=1}^T p(S_i | S_{i-1}; \theta),$$

where $p(S_0)$ denotes the probability distribution of a discrete random variable S_0 . In the study the latter does not constitute a quantity of interest, although it is straightforward to accommodate the sampling algorithm so that inference on S_0 is available.

Under the prior structure presented in Section 3, the following conditional posterior distributions are obtained:

$$- p(p_{ii} | \theta_{-p_{ii}}, h, S, y) = f_B(p_{ii} | a_i^*, b_i^*), \text{ for } i = 1, 2,$$

where

$$a_1^* = a_1 + n_{11}, \quad b_1^* = b_1 + n_{12},$$

$$a_2^* = a_2 + n_{22}, \quad b_2^* = b_2 + n_{21},$$

and

$$n_{ij} = \sum_{t=2}^T I(S_{t-1} = i)I(S_t = j);$$

$$- p(\sigma^2 | \theta_{-\sigma^2}, h, S, y) = f_{IG}(\sigma^2 | v_1^*, v_2^*),$$

where

$$v_1^* = \frac{T}{2} + v_1, \quad v_2^* = \left[\frac{1}{2} \sum_{i=1}^T (\ln h_i - m_i)^2 + \frac{1}{v_2} \right]^{-1};$$

$$- p(\beta | \theta_{-\beta}, h, S, y) = \begin{cases} f_N^{(3)}(\beta | \beta_*, \sigma^2 A_*^{-1}) I(\gamma_2 < 0) I(|\varphi| < 1) & \text{for MSSV } (\mu) \\ f_N^{(3)}(\beta | \beta_*, \sigma^2 A_*^{-1}) I(R_1 < 1) I(R_2 < 2) & \text{for MSSV } (\varphi) \end{cases}'$$

where

$$\beta_* = A_*^{-1} (\sigma^2 A_0 \beta_0 + W' \ln h), \quad A_* = \sigma^2 A_0 + W' W, \quad \ln h = (\ln h_1, \ln h_2, \dots, \ln h_T)',$$

and

$$W = \begin{bmatrix} 1 & I(S_1 = 2) & \ln h_0 \\ 1 & I(S_2 = 2) & \ln h_1 \\ \vdots & \vdots & \vdots \\ 1 & I(S_T = 2) & \ln h_{T-1} \end{bmatrix} \quad \text{for MSSV } (\mu),$$

$$W = \begin{bmatrix} 1 & I(S_1 = 1) \ln h_0 & I(S_1 = 2) \ln h_0 \\ 1 & I(S_2 = 1) \ln h_1 & I(S_2 = 2) \ln h_1 \\ \vdots & \vdots & \vdots \\ 1 & I(S_T = 1) \ln h_{T-1} & I(S_T = 2) \ln h_{T-1} \end{bmatrix} \quad \text{for MSSV } (\varphi);$$

$$p(h_i | \theta, h_{-i}, S, y) \propto \frac{1}{h_i^{1.5}} \exp\left(-\frac{y_i^2}{2h_i}\right) \exp\left(-\frac{1}{2\bar{\sigma}_i^2} (\ln h_i - w_i)^2\right),$$

where

for the MSSV(μ) model:

$$\text{for } t = 1, 2, \dots, T-1: \quad \bar{\sigma}_t^2 = \frac{\sigma^2}{1 + \varphi^2}, \quad w_t = \frac{\mu_{S_t} - \varphi \mu_{S_{t+1}} + \varphi (\ln h_{t+1} + \ln h_{t-1})}{1 + \varphi^2},$$

$$\text{for } t = T: \quad \bar{\sigma}_T^2 = \sigma^2, \quad w_T = \mu_{S_T} + \varphi \ln h_{T-1};$$

for the MSSV(φ) model:

$$\text{for } t = 1, 2, \dots, T-1: \quad \bar{\sigma}_t^2 = \frac{\sigma^2}{1 + \varphi_{S_{t+1}}^2}, \quad w_t = \frac{\mu(1 - \varphi_{S_{t+1}}) + \varphi_{S_t} \ln h_{t-1} + \varphi_{S_{t+1}} \ln h_{t+1}}{1 + \varphi_{S_{t+1}}^2},$$

$$\text{for } t = T: \quad \bar{\sigma}_T^2 = \sigma^2, \quad w_T = \mu + \varphi_{S_T} \ln h_{T-1}.$$

With regard to the conditional posterior distribution of vector S we only note here that it can be decomposed as:

$$p(S | \theta, h, y) = p(S_T | \theta, h, y) \prod_{t=1}^{T-1} p(S_t | S_{t+1}, \theta, h^t, y^t)$$

where y^t and h^t denotes the history of the observable process, and the volatility process, respectively, up to moment t . The idea has been suggested by Carter and Kohn (1994) and it is there that we refer for further details.

The MCMC sampling procedure

A full single cycle of the MCMC algorithm requires sampling each quantity of interest from its conditional posterior distribution. We employ the Gibbs procedure to sample the model parameters and vector S . For sampling the conditional volatilities, h_t 's, we adapt the Metropolis-Hastings algorithm used by Pajor (2003) in the case of non-switching SV models.

Let denote $\alpha^{(q)}$ the outcome on α from the q -th iteration³⁰ ($q = 1, 2, \dots, M, M+1, \dots, M+N$), and $\omega_{-\alpha}$ — a vector consisting of the elements of ω without its component α . Under this notation, a single full step of the sampling scheme proceeds as follows:

Step 1: sample $S^{(q+1)}$ from $p(S | \theta^{(q)}, h^{(q)}, S^{(q)}, y)$

Note: For a detailed description of the algorithm of sampling from $p(S | \theta, h, y)$ — see Carter and Kohn (1994);

Step 2: sample $p_{ii}^{(q+1)}$ from $p(p_{ii} | \theta_{-p_{ii}}^{(q)}, h^{(q)}, S^{(q+1)}, y)$, $i = 1, 2$;

Step 3: sample $\beta^{(q+1)}$ from $p(\beta | (\sigma^2)^{(q)}, p_{11}^{(q+1)}, p_{22}^{(q+1)}, h^{(q)}, S^{(q+1)}, y)$;

Step 3* — only for the MSSV(ϕ) model³¹ (*permutation sampler*; see Frühwirth-Schnatter, 2001):

→ if $\phi_1^{(q+1)} < \phi_2^{(q+1)}$ is violated, then:³²

³⁰ For $q = 0$ we obtain the set of starting values of the algorithm.

³¹ A note should be made here on sampling the model-specific parameters, β , when considering the MSSV(ϕ) model, i.e. Step 3* in the sampling scheme. In Section 2.2 the indentifiability constraint $\phi_1 < \phi_2$ is imposed. To guarantee that the restriction holds an additional step, called the *permutation sampler* (see Frühwirth-Schnatter, 2001), is introduced to the Gibbs procedure. Once a new β has been sampled from its full conditional posterior distribution, we check whether the inequality $\phi_1 < \phi_2$ is violated. If so, the subscripts '1' and '2' are simply interchanged so that the restriction is valid again. Since prior to sampling β all the state variables, S_t 's, and the transition probabilities are generated, they must be 'updated' (if ϕ_1 and ϕ_2 required switching), that is all the ones and twos in vector S as well as the subscripts of the probabilities p_{ij} need to be interchanged. Relevant assumptions, theorems and proofs of the validity of such an algorithm are found in Frühwirth-Schnatter (2001).

³² By ':' we denote the substitution operator.

$\varphi_1^{(q+1)} := \varphi_2^{(q+1)}$, $\varphi_2^{(q+1)} := \varphi_1^{(q+1)}$ (so that $\varphi_1^{(q+1)} < \varphi_2^{(q+1)}$ is guaranteed),

$S^{(q+1)} := 3l_T - S^{(q)}$ where $l_T = (1, 1, \dots, 1)_{(1 \times T)}$,

$p_{ij}^{(q+1)} := p_{3-i, 3-j}^{(q)}$ for $i, j = 1, 2$;

Step 4: sample $(\sigma^2)^{(q+1)}$ from $p(\sigma^2 | \theta_{-\sigma^2}^{(q+1)}, h^{(q)}, S^{(q+1)}, y)$;

Step 5: (the Metropolis-Hastings step): sample each $h_t^{(q+1)}$ from

$p(h_t | \theta^{(q+1)}, h_{(1:t-1)}^{(q+1)}, h_{(t+1:T)}^{(q)}, S^{(q+1)}, y)$, where $h_{(s:t)} = (h_s, h_{s+1}, \dots, h_t)$

with $s \leq t$.

MODEL STOCKA I WATSONA ORAZ JEGO MODYFIKACJE — ANALIZA INFLACJI W POLSCE

JACEK KWIATKOWSKI

Katedra Ekonometrii i Statystyki WNEIZ
Uniwersytet Mikołaja Kopernika w Toruniu
PL 87-100 Toruń, ul. Gagarina 13a
e-mail: jkwiat@uni.torun.pl

Praca przedstawiona na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie w dniu 30 kwietnia 2009 r.

ABSTRACT

Jacek Kwiatkowski. *Stock and Watson's Model and Its Modifications — Analysis of Inflation in Poland*, Folia Oeconomica Cracoviensia 2008–2009, 49–50: 145–168.

The paper presents general local level model with stochastic volatility, recently proposed for U.S. inflation by Stock and Watson. The main purpose is to present and compare other local level model specifications, especially with Normal GARCH and Student-t GARCH disturbances. The paper is a full Bayesian analysis and concerns inflation in Poland during 1992–2007. The model selection and posterior estimates provide strong evidence in favor of a model with heavy-tailed disturbances in the core component, and the transitory component. Also, after the system transformations in the early 90's, the volatility of the disturbances driving both components have been substantially decreasing over time.

KEY WORDS — SŁOWA KLUCZOWE

local level model, Bayesian Model Comparison, Inflation, conditional heteroscedasticity
Model lokalnego poziomu, bayesowskie testowanie modeli, inflacja,
warunkowa heteroskedastyczność

1. WSTĘP

Każdy poziom inflacji ma wpływ na procesy gospodarcze i społeczne. Wiele instytucji — rządowych, bankowych, finansowych, a także prywatni inwestorzy potrzebują wiarygodnego modelu inflacji aby móc przewidzieć prawdzi-

wą wartość bogactwa, dochodu oraz stopy zwrotu. Dlatego też tak ważne są badania mające na celu poprawne opisanie i prognozowanie tego zjawiska.

Problem modelowania inflacji jest znany od dawna zarówno w literaturze polskiej, jak i zagranicznej. Najnowszy przegląd w języku polskim tego typu literatury, głównie pod kątem analizy kointegracyjnej, przedstawia M. Majsterrek (2008).

W artykule przedstawiono nowe narzędzia służące do modelowania i prognozowania inflacji, mianowicie modele ze zmieniającymi się losowo w czasie parametrami. Modele te mogą stanowić ciekawą alternatywę w stosunku do znanych i szeroko stosowanych modeli strukturalnych (np. Welfe, 2000). W niedawnym raporcie dotyczącym inflacji w krajach G-7 Cecchetti, Hooper, Kasmann, Schoenholtz i Watson (2007) wykazali, że w ciągu ostatnich 50 lat wartość współczynnika autoregresji w modelu AR(1) dla szeregu inflacji zmieniała się znacząco w zakresie od $[-0,5$ do $0]$. Badania o podobnej tematyce podjęli wcześniej Koop i Potter (2001), którzy rozważali model autoregresyjny z parametrami generowanymi przez proces błędzenia przypadkowego dla kwartalnych danych amerykańskiego wskaźnika CPI (ang. *Consumer Price Index*). Przewaga tego modelu nad klasami konkurencyjnymi, w tym przypadku liniowym i progowym modelem autoregresyjnym, była tak duża, że model ten uzyskał prawdopodobieństwo *a posteriori* bliskie jeden¹. Co więcej, w obszernej analizie przeprowadzonej przez Stocka i Watsona (2007, 2008), które również dotyczyły inflacji w Stanach Zjednoczonych, żaden z modeli jednorodnaniowych, nie uzyskał dokładniejszej prognozy niż ich model². Rozważane przez nich konkurencyjne modele to zarówno typowe modele szeregów czasowych ARIMA, jak również modele oparte na krzywej Philipsa wraz z najnowszymi modyfikacjami oraz modele wykorzystujące koncepcję stopy bezrobocia nie powodującej przyspieszania inflacji — NAIRU. Przedstawiane w artykule modele inflacji są konstrukcjami formalnymi, które nie stawiają sobie za główny cel prawidłowe zdefiniowanie jej przyczyn. Główny nacisk położono na opis jej dynamiki i głębsze wniknięcie w jej wewnętrzną strukturę.

Rozpatrywane modele nawiązują do najnowszej literatury światowej, dotyczącej wykorzystania modeli z losowymi parametrami w analizie inflacji. O nowatorstwie prezentowanych w tym artykule tematyki może świadczyć, oprócz wspomnianej wcześniej literatury, również niepublikowany jeszcze artykuł Grassiego i Proiettiego (2008), w którym autorzy podejmują konkurencyjny, w stosunku do prezentowanego artykułu, kierunek badań.

Układ artykułu jest następujący. W części 2 przedstawiono model lokalnego poziomu, a następnie jego rozszerzenie w wersji Stocka i Watsona (2007). W dalszej części artykułu zaproponowano modyfikacje wspomnianego mode-

¹ Mowa tu o modelach TAR (ang. *Threshold Autoregressive Models*).

² Wg kryterium pierwiastka błędu średniokwadratowego (RMSE).

lu, które mają na celu, po pierwsze, w większym stopniu uwzględnić typowe własności szeregu inflacji, po drugie — mogą poprawić jego możliwości aplikacyjne. Dotyczy to zwłaszcza wykorzystania modelu GARCH z warunkowym rozkładem normalnym i *t*-Studenta w równaniu obserwacji i nieobserwowalnej zmiennej — trendu. Dodatkowo omówiono inne konkurencyjne specyfikacje przydatne i stosowane wcześniej do modelowania inflacji, to jest modele autoregresyjne z stałymi oraz losowymi parametrami. W tym drugim przypadku przyjmuje się założenie, że parametry opisane są jako proces błędzenia przypadkowego. W części 3 zbadano własności wymienionych modeli na podstawie 192-miesięcznych obserwacji polskiego wskaźnika cen konsumenta CPI, obejmujących okres od stycznia 1992 do grudnia 2007 roku. W tym celu na wstępie zbadano rząd integracji wskaźnika CPI, a w dalszej kolejności dokonano bayesowskiej estymacji modeli, a także porównano ich moc objaśniającą. Zbadano również zmienność indeksu cen oraz długookresowego trendu poziomu cen, a także obliczono jak kształtował się w badanym okresie współczynnik korelacji między przyrostami CPI a ich pierwszymi opóźnieniami. Część 4 zawiera wnioski.

2. MODEL STOCKA I WATSONA ORAZ JEGO MODYFIKACJE

Modelowanie inflacji jest jednym z podstawowych zagadnień współczesnej makroekonometrii. Problemem modelowania inflacji w Polsce jest szeroko omawiany, m.in. w książkach z zakresu ekonometrii: Welfe (1993), Osińska (2000), Kołowski (2006) oraz Majsterek (2008). W bieżącej części proponuje się inne niż w przytoczonej literaturze narzędzie służące do opisu inflacji — model Stocka i Watsona, będący szczególną specyfikacją dobrze znanego w literaturze modelu lokalnego poziomu³.

Do modelowania inflacji Stock i Watson (2007) użyli zmodyfikowany model lokalnego poziomu — LL-SV (ang. *Local Level Model*):

$$y_t = \delta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad (2.1)$$

$$\delta_t = \delta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \omega_t^2), \quad (2.2)$$

gdzie przez y_t dla $t = 1, \dots, T$ oznaczono obserwacje zmiennej zależnej, natomiast δ_t jest bieżącym, nieobserwowalnym bezpośrednio poziomem procesu w czasie t , z kolei ε_t jest białym szumem w równaniu obserwacji. Przyjmuje się, że zmiana w czasie poziomu obserwowanego procesu odbywa się według prostego procesu błędzenia przypadkowego z błędem η_t . Warunkowe wariacje są opisane poprzez procesy stochastycznej zmienności:

³ W pracy model lokalnego poziomu z warunkowymi wariacjami opisywanymi przez model zmienności stochastycznej oznaczono jako LL-SV, natomiast w artykułach Stocka i Watsona (2007, 2008) występuje on jako UC-SV (ang. *Unobserved Component*).

$$h_{szum,t} = h_{szum,t-1} + \xi_{szum,t} \quad (2.3)$$

$$h_{sygnat,t} = h_{sygnat,t-1} + \xi_{sygnat,t} \quad (2.4)$$

gdzie:

$$\sigma_t^2 = \exp(h_{szum,t}) \text{ i } \omega_t^2 \exp(h_{sygnat,t}) \text{ oraz } \xi_{szum(sygnat)} \sim N(0, y^2).$$

Model ten zakłada, że pierwsze przyrosty Δy_t mają zmienną w czasie i ujemną, warunkową autokorelację rzędu pierwszego i brak autokorelacji wyższych rzędów. Różnicując (2.1) mamy:

$$\Delta y_t = \eta_t = \varepsilon_t - \varepsilon_{t-1}. \quad (2.5)$$

Przy danym parametrze γ^2 , wariancja pierwszych przyrostów jest warunkowa względem zmiennych ukrytych $h_{szum,t}$ i $h_{sygnat,t}$ oraz całej przeszłości Ψ_{t-1} i jest równa:

$$\text{Var}(\Delta y_t | \Psi_{t-1}, h_{szum,t}, h_{sygnat,t}, \gamma^2) = \omega_t^2 + 2\sigma_t^2. \quad (2.6)$$

Natomiast warunkowa kowariancja między Δy_t i Δy_{t-1} wynosi:

$$\text{cov}(\Delta y_t, \Delta y_{t-1} | \Psi_{t-1}, h_{szum,t}, h_{sygnat,t}, \gamma^2) = -\sigma_t^2. \quad (2.7)$$

Ponieważ wariancja jest zawsze dodatnia, warunkowy współczynnik autokorelacji przyjmuje zawsze wartość ujemną i jest zmienny w czasie:

$$\text{corr}(\Delta y_t, \Delta y_{t-1} | \Psi_{t-1}, h_{szum,t}, h_{sygnat,t}, \gamma^2) = \frac{-\sigma_t^2}{\omega_t^2 + 2\sigma_t^2}. \quad (2.8)$$

Kiedy wariancje σ_t^2 i ω_t^2 są stałe, model (2.1)–(2.4), redukuje się do tradycyjnego modelu lokalnego poziomu, który dla pierwszych przyrostów ma taką samą strukturę jak model IMA(1,1).

Zaproponowany przez Stocka i Watsona model nie uwzględnia jednak kilku istotnych własności analizowanych szeregów czasowych. Po pierwsze, w równaniach (2.3) i (2.4) współczynnik autoregresji jest równy jeden, co oznacza, że warunkowa wariancja podlega błędzeniu przypadkowemu. Implikuje to także niestacjonarność w sensie kowariancji pierwszych różnic, tj. logarytmicznych przyrostów wskaźnika cen konsumenta:

$$\text{Var}(\Delta y_t | \Psi_{t-1}, h_{szum,t}, h_{sygnat,t}, \gamma^2) = \omega_t^2 + 2\sigma_t^2 \rightarrow \infty \text{ dla } t \rightarrow \infty,$$

co z kolei oznacza, że jest to proces I(2), a nie I(1). Po drugie, występuje jedna, wspólna wariancja resztowa γ^2 w równaniach warunkowych wariancji (2.3) i (2.4).

Istnieje zatem potrzeba stworzenia bardziej ogólnej specyfikacji modelu Stocka i Watsona, która uwzględniałaby w szerszym stopniu własności szeregu inflacji. Ciekawe byłyby również badania polegające na zastąpieniu w opi-

się warunkowej wariancji modelu stochastycznej zmienności, modelem GARCH. Modele GARCH to, jak dotąd, modele najczęściej stosowane w opisie zmienności szeregów (Doman i Doman, 2004), dlatego też warto sprawdzić ich przydatność w modelowaniu zmienności wskaźnika inflacji. Z tego też względu rozpatrzono kilka konkurencyjnych i wzajemnie wykluczających się modyfikacji modelu Stocka i Watsona.

Pierwszy analizowany model to model lokalnego poziomu z błędami typu SV, którego postać jest mocno zbliżona do modelu w wersji zaproponowanej przez Stocka i Watsona. Następne dwa modele to modele lokalnego poziomu, w których wariancja warunkowa w równaniu obserwacji i trendu jest opisana jako GARCH(1,1). Dla modeli GARCH rozpatrzono dwa warianty, pierwszy i zarazem prostszy, to model, w którym reszty (w równaniu obserwacji i trendu) przyjmują warunkowe rozkłady normalne, oraz drugi — w którym założono grubsze ogony poprzez przyjęcie warunkowych rozkładów *t*-Studenta. Własności modelu lokalnego poziomu, w którym warunkowe wariancje w równaniu obserwacji i nieobserwowalnego trendu są procesem GARCH(1,1), opisują Pellegrini, Ruiz i Espasa (2007, 2008). Kolejny, czwarty — to model autoregresyjny z parametrami generowanymi przez procesy błędzenia przypadkowego z warunkową wariacją w równaniu obserwacji typu SV. Model ten oznaczono jako RCA-SV (RW). Podobny model tylko, że z homoskedastyczną wariancją użyli Koop i Potter (2001) dla szeregu inflacji w Stanach Zjednoczonych. Ostatni, piąty model jest najprostszy, ponieważ jest to standardowy model autoregresyjny (rzędu drugiego) również z błędami SV. Jest on zapisany jako AR(2)-SV.

Jeżeli założymy warunkową normalność procesów resztowych to wariancje w równaniu obserwacji i trendu mogą przyjąć następującą postać:

$$\varepsilon_t \sim N(0, \sigma_t^2) \text{ i } \eta_t \sim N(0, \omega_t^2), \quad (2.9)$$

gdzie:

$$\sigma_t^2 = h_{szum,t} \text{ i } \omega_t^2 = h_{sygnal,t} \quad (2.10)$$

$$h_{szum,t} = b_{1szum} h_{szum,t-1} + a_{0,szum} + a_{1,szum} (E\varepsilon_{t-1})^2, \quad (2.11)$$

$$h_{sygnal,t} = b_{1sygnal} h_{sygnal,t-1} + a_{0,sygnal} + a_{1,sygnal} (E\eta_{t-1})^2, \quad (2.12)$$

wraz z standardowymi założeniami zapewniającymi dodatniość i skończoność warunkowych wariancji:

$$a_{0,szum(sygnal)} \equiv 1 - b_{1,szum(sygnal)} - a_{1,szum(sygnal)}, b_{1,szum(sygnal)} \geq 0,$$

$$a_{1,szum(sygnal)} \geq 0, a_{1,szum(sygnal)} + b_{1,szum(sygnal)} < 1.$$

Z postaci warunkowych wariancji wynika, że mamy: $\text{var}(\varepsilon_t) = 1$ i $\text{var}(\eta_t) = 1$ oraz $E(h_{szum(sygnal),t}) = 1$.

Równania (2.10)–(2.12) opisują modele GARCH(1,1) z warunkowym rozkładem normalnym. Model lokalnego poziomu z tego typu błędami będzie w dalszej części artykułu zapisywany w skrócie jako LL-GARCH. Jako alternatywę można również użyć zamiast warunkowego rozkładu normalnego rozkład t -Studenta. Odpowiednie reszty z rozkładem t -Studenta oznaczono w następujący sposób:

$$\varepsilon_t \sim t\left(0, \frac{v_{szum}}{(v_{szum} - 2)\sigma_t^2} v_{szum}\right) \text{ i } \eta_t \sim t\left(0, \frac{v_{sygnat}}{(v_{sygnat} - 2)\omega_t^2} v_{sygnat}\right), \quad (2.13)$$

gdzie $t(a, P, v)$ oznacza jednowymiarowy rozkład Studenta o niecentralności a , precyzji P i stopniach swobody v . Model ten oznaczono w skrócie jako LL-GARCH-Student.

Kolejnym możliwym wariantem postaci wariancji procesów resztowych jest model zmienności stochastycznej. Jest on oznaczony jako LL-SV. W tym przypadku wariancje zmieniają się w czasie zgodnie z:

$$h_{szum,t} = \rho_{szum} h_{szum,t-1} + \zeta_{szum,t}, \quad (2.14)$$

$$h_{sygnat,t} = \rho_{sygnat} h_{sygnat,t-1} + \zeta_{sygnat,t}, \quad (2.15)$$

gdzie: $\varepsilon_t \sim N(0, \sigma_t^2)$ i $\eta_t \sim N(0, \omega_t^2)$, $\sigma_t^2 = \exp(h_{szum,t})$ i $\omega_t^2 = \exp(h_{sygnat,t})$ oraz $\rho_{szum(sygnat)} \in (-1, 1)$, a także $\xi_{szum(sygnat)} \sim N(0, \gamma_{szum(sygnat)}^2)$.

Proces ten w odróżnieniu od procesu Stocka i Watsona (2007), poprzez nałożenie warunków ograniczających na współczynniki autoregresji, ma stacjonarne warunkowe wariancje, a co zatem idzie jest I(1), a nie I(2). Dodatkowo, dzięki dwóm wariancjom resztowym w równaniach dla h_t można rozpatrywać różną zmienność warunkowych wariancji w równaniu obserwacji i trendu.

Model autoregresyjny rzędu drugiego, w którym zarówno stała, jak i współczynniki autoregresji generowane są przez proces błędzenia przypadkowego ma następującą postać:

$$y_t = \delta_{0t} + \delta_{1t} y_{t-1} + \delta_{2t} y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad (2.16)$$

$$\delta_{it} = \delta_{i,t-1} + \eta_{it}, \quad \eta_{it} \sim N(0, \omega_t^2) \text{ dla } i = 0, 1, 2, \quad (2.17)$$

gdzie: $\varepsilon_t \sim N(0, \sigma_t^2)$, $\sigma_t^2 = \exp(h_{szum,t})$ oraz $h_{szum,t} = \rho_{szum} h_{szum,t-1} + \zeta_{szum,t}$, $\rho_{szum} \in (-1, 1)$, a także $\xi_{szum} \sim N(0, \gamma_{szum}^2)$.

Nakładając warunki ograniczające na wariancje $\omega_t^2 = 0$ dla $i = 0, 1, 2$, w modelu (2.16)–(2.17), uzyskuje się standardowy model AR(2)-SV:

$$y_t = \delta_{const} + \delta_{1,const} y_{t-1} + \delta_{2,const} y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2). \quad (2.18)$$

Wszystkie przedstawione powyżej modele zostały wykorzystane w następnej części artykułu do analizy wskaźnika inflacji w Polsce.

3. ANALIZA INDEKSU CEN KONSUMENTA W POLSCE W LATACH 1992–2007

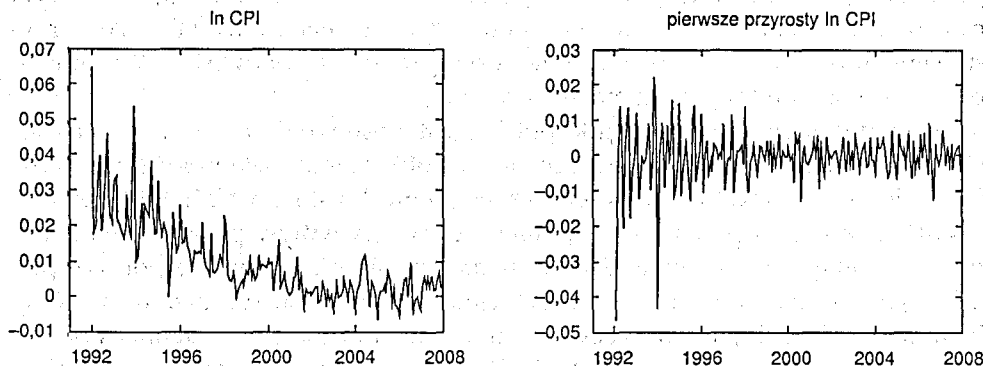
W tym punkcie przeprowadzono analizę szeregu inflacji w Polsce przy użyciu, omówionych przed chwilą modeli z rodziny LL. Celem analizy było zestawienie podobieństw i różnic w modelowaniu wskaźnika cen konsumenta, z użyciem konkurujących i wzajemnie wykluczających się modeli, a także porównanie ich mocy objaśniającej. Na początku wszystkie modele poddano estymacji, a następnie obliczono ich moc objaśniającą. Porównanie stopnia dopasowania poszczególnych modeli, dokonano standardowo, za pomocą brzegowej gęstości wektora obserwacji, która była obliczona za pomocą metody Newtona-Raferty'ego. Inflacja użyta w niniejszym artykule mierzona jest jako logarytm wskaźnika cen dóbr i usług konsumpcyjnych, w którym za podstawą wzięto okres wcześniejszy. Dane wykorzystywane do estymacji modeli i testowania ich mocy objaśniającej dotyczą okresu od stycznia 1992 do grudnia 2007 roku. W ten sposób otrzymano 192 obserwacje miesięczne. Wybór początkowego roku, od którego rozpoczęto badania, był uzależniony z jednej strony chęcią uzyskania możliwie jak najdłuższego szeregu czasowego, z drugiej strony zaś podlegał ograniczeniom, ponieważ pominięto okres hiperinflacji, która była następstwem głębokich przemian gospodarczych w latach 1989/90. Przed przystąpieniem do analizy badany szereg oczyszczono z sezonowości. W tym celu posłużono się średnią ruchomą scentrowaną⁴.

Badany okres, jak na warunki polskie jest stosunkowo długi. Jest to okres, który obejmuje zarówno głębokie zmiany w polityce makroekonomicznej, transformację gospodarki, jak również przystąpienie Polski do Unii Europejskiej. Na długookresowy poziom inflacji ma z pewnością wpływ prowadzona przez bank centralny polityka pieniężna. W badanym okresie zmiany w polityce pieniężnej miały charakter zarówno instytucjonalny, np. ukonstytuowanie się Rady Polityki Pieniężnej, jak i dotyczyły ewolucji w zakresie prowadzonej polityki pieniężnej w tym jej celów i strategii. W przeprowadzonej analizie pominięto nietypowy, początkowy okres głębokiej transformacji systemowej, kiedy miały miejsce zjawiska hiperinflacyjne a inflacja sięgała nawet tysiąc procent w skali roku. Mimo to pierwszy rok badanego szeregu (1992) przypada jeszcze na koniec głębokiej transformacji systemowej, co daje możliwość uzyskania ciekawych wniosków. Z kolei następne lata (1993–1997) to głównie trwałe i stopniowe przeprowadzanie procesu dezinflacji, będącej skutkiem wprowadzania planu

⁴ Do eliminacji sezonowości w szeregu inflacji wykorzystano metodę średniej ruchomej w programie Eviews 6 z następującymi współczynnikami: 0,007589; -0,000333; -0,000518; 0,002364; -0,000284; -0,002781; -0,008910; -0,005796; 0,005922; 0,001802; 0,000217; 0,000728. Inne metody oczyszczania szeregu z sezonowości takie jak Census X12 i X11 dały bardzo zbliżone wyniki i nie wpływały znacząco na charakter oczyszczonego szeregu.

stabilizacyjnego. Kolejne lata (1998–2004) to większa niezależność banku centralnego oraz okres przygotowania Polski do integracji ze strukturami Unii Europejskiej i, co za tym idzie, istotne zmiany w polityce pieniężnej i antyinflacyjnej: m.in. przyjęcie nowej strategii monetarnej polegającej na wyznaczeniu średniookresowej strategii polityki pieniężnej oraz na ustalaniu bezpośredniego celu inflacyjnego. Na osiągnięcie inflacji na poziomie 4–5% potrzebowano około dwunastu lat, co zdaniem niektórych ekonomistów (np. Grabia, 2003) świadczy jednak o zbyt powolnym tempie wygaszania tego niekorzystnego zjawiska. Po roku 2004 głównym celem polityki pieniężnej stało się stabilizowanie inflacji na niskim poziomie oraz dążenie do spełnienia kryteriów zapisanych w Traktacie z Maastricht⁵.

Indeks cen konsumenta (CPI) oraz jego pierwsze przyrosty, po wcześniejszym logarytmowaniu i oczyszczeniu z sezonowości, przedstawia rycina 1. Patrząc na przebieg inflacji, można dojść do kilku ciekawych wniosków. Po pierwsze, wyraźnie widać, że badany okres można podzielić na dwa zasadnicze podokresy: lata od 1992 do 2003, kiedy nastąpił wyraźny spadek poziomu inflacji oraz okres po roku 2003, w którym inflacja oscylowała wokół względnego stałego poziomu.



Źródło: obliczenia własne na podstawie danych publikowanych przez GUS, <http://www.stat.gov.pl>

Ryc. 1. Logarytm wskaźnika CPI (po wyeliminowaniu sezonowości) oraz jego pierwsze różnice w okresie od stycznia 1992 do grudnia 2007 roku

Po drugie, biorąc pod uwagę pierwsze przyrosty wskaźnika CPI, można zauważyć, że lata wcześniejsze, szczególnie między 1992 a 1997 charakteryzują się znacznie większą zmiennością niż lata późniejsze.

Model Stocka i Watsona daje możliwość opisu tych wymienionych własności, tj. zmiany poziomu procesu i efektu skupiania zmienności (ang. *volatility*

⁵ Przegląd i ocenę polityki pieniężnej od roku 1989 dokonuje np. Przybylska-Kapuścińska, 2006.

clustering). Efekt skupiania zmienności jest typową własnością szeregów finansowych (patrz Doman i Doman, 2004) i polega na tym, że zarówno małe, jak i duże zmiany cen aktywów następują seriami. Najnowsze badania empiryczne (np. Cecchetti i in., 2007) potwierdzają występowanie efektu skupiania zmienności, a co za tym idzie — zmiennej w czasie warunkowej wariancji w szeregu inflacji. Model LL posiada dodatkowo jeszcze jedną ciekawą własność, której nie można w prosty sposób zauważyć na rycinie 1, mianowicie poprzez warunkową wariancję w równaniu stanu zakłada, że mogą występować znaczne wahania długookresowego poziomu cen.

Na wstępie analizy przystąpiono do zbadania stopnia integracji danych dotyczących wskaźnika cen konsumenta CPI. W tym celu zastosowano bayesowski test pierwiastka jednostkowego, który zaproponowali Koop i van Dijk (2000). Analizą stopnia integracji, w porównywalnym okresie, dla tego samego szeregu, z użyciem testu Dickeya-Fullera i KPSS⁶, zajmowali się m.in. Welfe i Majsterek (2002).

Rozważmy model, który obrazuje związki między popularnymi w literaturze klasycznej (niebayesowskiej) testami pierwiastka jednostkowego, czyli Dickeya-Fullera i KPSS:

$$\ln CPI_t = \tau_t + \rho \ln CPI_t + \sum_{i=1}^{p-1} c_i \Delta \ln CPI_{t-i} + v_t, \quad v_t \sim N(0, \sigma_v^2), \quad (3.1)$$

$$\tau_t = \alpha + \tau_{t-1} + u_t, \quad u_t \sim N(0, \sigma_u^2). \quad (3.2)$$

Wykorzystując prostą transformację $\lambda = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$, jednostajny rozkład *a priori* $p(\lambda) = 1$ w przedziale $[0, 1)$ oraz rząd autoregresji $p = 2$, możemy rozważyć następujące cztery rozłączne hipotezy⁷:

H_1 : $\lambda = 0$ i $|\rho| < 1$. Szereg jest stacjonarny wokół deterministycznego trendu.

H_2 : $0 < \lambda < 1$ i $|\rho| = 1$. Szereg jest ARIMA z dryfem.

H_3 : $\lambda = 0$ i $|\rho| < 1$. Szereg jest I(1) wraz z deterministycznym trendem.

H_4 : $0 < \lambda < 1$ i $|\rho| = 1$. Szereg jest zintegrowany rzędu drugiego, I(2).

Wyniki testowania rzędu integracji dla logarytmu indeksu inflacji są zawarte w tabeli 1⁸. Jak widać wszystkie hipotezy miały jednakowe prawdopodobieństwo *a priori*. Trzeci wiersz w tabeli 1 pokazuje wartości brzegowej gęstości obserwacji obliczone w modelach, które były powiązane z jedną z czterech hipotez. Modele te uzyskano poprzez nałożenie warunków ograniczających zgodnie z wymienionymi wyżej hipotezami. Jak widać, znaczną przewagę uzyskała hipoteza, która głosi, że pierwsze przyrosty badanego szeregu są sta-

⁶ Dickey i Fuller, 1979, Kwiatkowski i in., 1992.

⁷ Koop i van Dijk, 2000, wyrażenie (A.3).

⁸ Ze względu na ograniczony zakres jakie mogą przyjmować wartości rzeczywiste w języku programowania Matlab, w każdej iteracji od logarytmu funkcji wagowej w metodzie Monte Carlo z funkcją ważności odejmowano 700.

Prawdopodobieństwa a posteriori hipotez dotyczących stopnia integracji, obliczone dla logarytmu indeksu inflacji

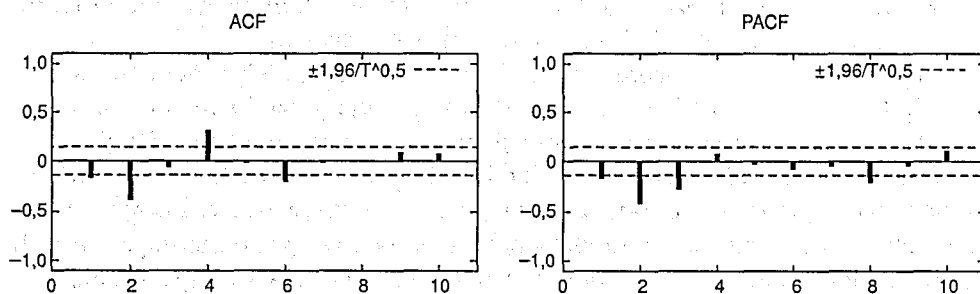
	$P(H_1 \text{Dane})$	$P(H_2 \text{Dane})$	$P(H_3 \text{Dane})$	$P(H_4 \text{Dane})$
Prawdopodobieństwo a priori	0,25	0,25	0,25	0,25
Brzegowa gęstość obserwacji	1,78E + 100	1,51E + 118	1,78E + 80	2,93E + 77

Źródło: obliczenia własne.

cjonarne. Wartość brzegowej gęstości obserwacji dla H_2 jest o osiemnaście rzędów większa niż dla H_1 i około półtora razy większa niż dla pozostałych dwóch hipotez. Z tego też względu, w tabeli 1 świadomie zrezygnowano z podania prawdopodobieństw a posteriori, ponieważ z prawdopodobieństwem bliskim wartości jeden można przyjąć hipotezę o stacjonarności pierwszych przyrostów.

Wyniki zawarte w tabeli 1 są zgodne z wynikami, jakie uzyskali Welfe i Majsterek (2002) w analizie stopnia integracji cen dla miesięcznych obserwacji wskaźnika CPI i są kolejnym argumentem w prowadzonej dyskusji (por. Majsterek, 2008) za traktowaniem inflacji w Polsce jako $I(1)$, a co za tym idzie — za przyjęciem zintegrowania rzędu drugiego w procesie generującym ceny.

Jak wiadomo, modele lokalnego poziomu implikują ujemną autokorelację pierwszych przyrostów. Warto więc sprawdzić, czy własność ta występuje w szeregu inflacji. W tym celu obliczono wartości funkcji autokorelacji (ACF) i autokorelacji cząstkowej (PACF). Wyniki obliczeń wraz z przerywaną linią, która oznacza typowy 95% przedział ufności, przedstawia rycina 2.



Źródło: obliczenia własne.

Ryc. 2. Funkcja autokorelacji (ACF) i autokorelacji cząstkowej (PACF) wraz z przedziałami ufności obliczona dla przyrostów logarytmów CPI

Jak widać, cztery z dziesięciu wartości ACF są istotne statystycznie przy 5-procentowym poziomie istotności, z czego pierwsza, druga i szósta przyjmują wartości ujemne. Wyniki te potwierdzają hipotezę o istnieniu ujemnej auto-

korelacji dla pierwszych przyrostów inflacji. O ile ujemna autokorelacja nie budzi wątpliwości, to sugerowany, na podstawie oszacowanych wartości PACE, rząd autokorelacji jest niejednoznaczny. Najbardziej istotne okazało się drugie opóźnienie, potem w kolejności malejącej: trzecie, ósme i pierwsze. Sugeruje to istotną autokorelację rzędu drugiego lub trzeciego. Postać funkcji autokorelacji, przedstawiona na rycinie 2, jest zasadniczo zgodna z postacią funkcji autokorelacji obliczoną dla szeregów inflacji krajów grupy G-7, gdzie jak podają Cecchetti i in. (2007), również uzyskano istotną ujemną autokorelację rzędu pierwszego⁹. Pewne wątpliwości budzi jednak istotność autokorelacji nie tylko rzędu pierwszego, ale również wyższych rzędów, a także wartości funkcji autokorelacji cząstkowej, co wskazuje na bardziej złożoną strukturę niż w modelu IMA(1,1). Może to wpływać w tym przypadku na moc objaśniającą modeli lokalnego poziomu, ponieważ zakładamy w nich *explicite* występowanie ujemnej, istotnej autokorelacji rzędu pierwszego i brak autokorelacji wyższych rzędów oraz ujemne wartości autokorelacji cząstkowej, dążące do zera w tempie wykładniczym.

Wnioskowanie bayesowskie wymaga przyjęcia rozkładów *a priori* estymowanych parametrów. W tym przypadku wymiar wektora parametrów oraz jego współrzędne różnią się znacznie, w zależności od rozpatrywanej hipotezy modelowej. Szczegółowe informacje na temat rozkładów *a priori* szacowanych parametrów, w badanych modelach, przedstawia tabela 2. Jak widać, są to rozkłady właściwe, w większości bardzo rozproszone, odzwierciedlające nieprecyzyjną wiedzę badacza na temat ich prawdziwych wartości. Dodatkowo, zmienna odpowiadająca za trend — δ_t oraz za warunkową wariancję w modelach SV — h_t , a także zmienne będące współczynnikami autoregresji w modelach RCA-SV (RW) — δ_{it} traktowane są jak zmienne ukryte (ang. *latent variable*), czyli zmienne nie dające się bezpośrednio obserwować. W podejściu bayesowskim estymacji zmiennych ukrytych dokonuje się w taki sam sposób, jak estymacji pozostałych parametrów modelu, a ich rozkład *a priori* wynika bezpośrednio z hierarchicznej struktury rozpatrywanych modeli (Koop, 2003). Szczegóły bayesowskiej estymacji modelu lokalnego poziomu o rozkładach dopuszczających warunkowy rozkład *t*-Studenta i zmienną wariancję omówiono w pracy Kwiatkowskiego (2009). W literaturze polskiej o bayesowskiej estymacji zmiennych ukrytych, głównie w kontekście modeli stochastycznej zmienności, pisze Pajor (2003). Inny ciekawy przykład zastosowania zmiennych ukrytych, tym razem dla funkcji produkcji Cobba i Douglasa, przedstawia Osiewalski (2001).

Dla początkowego stanu δ_0 , w modelach lokalnego poziomu, przyjęto wartość równą 6, natomiast w modelu RCA-SV (RW) dla początkowych wartości

⁹ Prezentowane wyniki najbardziej zbliżone są do wyników otrzymanych dla szeregu inflacji w Kanadzie w latach 1984–2006, por. Cecchetti i in., 2007, tablica 3.1, s. 14.

Rozkłady *a priori* parametrów

Parametr	Przyjęty rozkład <i>a priori</i>	Hiperparametry	Użyty w modelu
$a_{1, szum(sygnal)}$	Jednostajny w obszarze stacjonarności i dodatniości warunkowej wariancji		LL-GARCH, LL-GARCH-Student
$b_{1, szum(sygnal)}$	Jednostajny w obszarze stacjonarności i dodatniości warunkowej wariancji		LL-GARCH, LL-GARCH-Student
$\nu_{szum(sygnal)}$	Ucięty Cauchy'ego, $\nu_{szum(sygnal)} > 2$		LL-GARCH-Student
$\rho_{szum(sygnal)}$	Ucięty normalny $N(\mu_\rho, \sigma_\rho^2), \rho_{szum(sygnal)} \in (-1, 1)$	$\mu_\rho = 0,95,$ $\sigma_\rho^2 = 0,5$	LL-SV, RCA(2)-SV (RW), AR(2)-SV
$\gamma_{szum(sygnal)}^2$	Odwrócony gamma $IG(c, d), \gamma_{szum(sygnal)}^2, \sigma_i^2, \omega_i^2 > 0$	$c = 0,01,$ $d = 0,01$	LL-SV, RCA(2)-SV (RW), AR(2)-SV
σ^2			RCA(2)-SV (RW), AR(2)-SV
ω_i^2 , dla $i = 0, 1, 2.$			RCA(2)-SV (RW)
δ_{const}	Normalny $N(\mu_\delta, \sigma_\delta^2)$	$\mu_\delta = 0,$ $\sigma_\delta^2 = 1$	AR(2)-SV
$\delta_{i, const}$ dla $i = 1, 2.$	Ucięty normalny w obszarze stacjonarności $N(\mu_\delta, \sigma_\delta^2)$		AR(2)-SV

współczynników autoregresji oraz stałej: δ_{10} , δ_{20} i δ_{00} , założono rozkład normalny, o średniej i wariancji równej odpowiednio 0 i 0,1. Wartość początkowa warunkowej wariancji jest traktowana jako ustalona i równa $h_0 = 1$ w modelu GARCH, i $h_0 = 0$ w modelu SV.

Dla wszystkich wariancji, zarówno w standardowych modelach AR, jak i w modelach autoregresyjnych z losowymi parametrami oraz zmienności stochastycznej, przyjęto odwrócony rozkład gamma. Dla parametrów w modelu GARCH, stojących przy opóźnionych kwadratach reszt i opóźnionej wariancji, przyjęto rozkłady jednostajne na sympleksach otrzymanych poprzez restrykcje gwarantujące dodatniość i skończoność procesu (por. Osiewalski i Pipień, 1999). Dla współczynników autoregresji, w standardowym modelu AR(2), założono ucięty wielowymiarowy rozkład normalny o zerowym wektorze średnich i diagonalnej macierzy kowariancji, z elementami na głównej przekątnej równymi wartości jeden. Również dla stałej δ_{const} w modelu AR(2), przyjęto rozkład normalny o średniej równej zero i wariancji jeden. Z kolei dla współczynnika au-

toregresji, w modelu SV, wykorzystano wstępną wiedzę na jego temat i przyjęto stosunkowo informacyjny rozkład normalny ze średnią i wariancją równą odpowiednio 0,95 i 0,5¹⁰. Dla liczby stopni swobody przyjęto rozkład Cauchy'ego, ucięty lewostronnie w $v = 2$. Jest to zatem rozkład *a priori* w którym z jednej strony założono istnienie warunkowej wariancji procesów (w równaniu obserwacji i trendu), z drugiej strony przyjęto taką masę prawdopodobieństwa, aby grube ogony rozkładu warunkowego miały najwyższą gęstość prawdopodobieństwa. Dodatkowo, z własności rozkładu Cauchy'ego wiadomo, że rozkład ten nie posiada skończonych momentów. Oznacza to, że zarówno wariancja jak i kurtoza nie istnieją. Własność ta, wraz z konstrukcją przedziału, który jest prawostronnie nieograniczony, dopuszcza realną możliwość występowania cieńszych ogonów, poprzez istnienie warunkowej normalności procesów.

W tabelach 3 i 4 zamieszczono rezultaty bayesowskiej estymacji parametrów w modelach opisujących logarytmny CPI lub ich przyrosty. Przed przystą-

Tabela 3

Informacje *a posteriori* parametrów, w modelach LL-GARCH i LL-GARCH-Student, obliczone dla miesięcznych obserwacji wskaźnika CPI w latach 1992–2007

Parametr		LL-GARCH		LL-GARCH-Student
$a_{1, \text{sygnal}}$	0,2597	(0,1345)	0,2775	(0,1331)
	0,2238	[0,0701 0,5501]	0,2578	[0,0557 0,5388]
$b_{1, \text{sygnal}}$	0,7286	(0,1477)	0,7007	(0,1503)
	0,7700	[0,4170 0,9406]	0,7256	[0,4064 0,9424]
v_{sygnal}	—	—	2,4891	(0,4135)
			2,3890	[2,0827 3,0946]
$a_{1, \text{szum}}$	0,2456	(0,1458)	0,2713	(0,1637)
	0,2083	[0,0695 0,5243]	0,2373	[0,0369 0,6007]
$b_{1, \text{szum}}$	0,7431	(0,1594)	0,6090	(0,2563)
	0,7857	[0,4400 0,9304]	0,6802	[0,0935 0,9454]
v_{szum}	—	—	2,4796	(0,4106)
			2,3831	[2,1173 3,0580]
$a_{1, \text{sygnal}} + b_{1, \text{sygnal}}$	0,9919	(0,0099)	0,9783	0,0241
	0,9952	[0,9725 1,0000]	0,9852	[0,9310 0,9998]
$a_{1, \text{szum}} + b_{1, \text{szum}}$	0,9931	(0,0087)	0,8668	(0,1475)
	0,9959	[0,9768 1,0000]	0,9266	[0,5536 0,9988]

Źródło: obliczenia własne.

¹⁰ Na przykład, dla kwartalnych danych wskaźnika konsumenta, w artykule Grassiego i Proiettiego (2008) wartość oczekiwana rozkładu *a posteriori* współczynników autoregresji, w modelach SV, wyniosła odpowiednio 0,9883 w równaniu obserwacji i 0,9856 w równaniu trendu.

Informacje a posteriori parametrów, w modelach LL-SV i RCA(2)-SV (RW), obliczone dla miesięcznych obserwacji wskaźnika CPI w latach 1992-2007

Parametr		LL-SV	Parametr		RCA(2)-SV (RW)
ρ_{sygnal}	0,9930 0,9950	(6,1735e-3) [0,9837 0,9990]	ρ_{szum}	0,9780 0,9899	(0,0332) [0,9129 0,9990]
γ^2_{sygnal}	0,1816 0,1255	(0,1815) [0,0098 0,5224]	γ^2_{szum}	0,46351e-1 0,30729e-1	(0,58458e-1) [4,1758e-3 1,3339e-1]
			σ^2	0,7817 0,6572	(0,5558) [0,1444 1,8100]
ρ_{szum}	0,9906 0,9932	(8,7786e-3) [0,9775 0,9990]	ω_0^2	2,8213e-3 2,5071e-3	(1,3714e-3) [9,7113e-4 5,5280e-3]
γ^2_{szum}	0,6404 0,3546	(1,0457) [0,0260 1,9919]	ω_1^2	5,1339e-3 4,3482e-3	(3,1228e-3) [1,0952e-3 1,1065e-2]
			ω_2^2	6,4763e-3 5,4202e-3	(4,1621e-3) [1,5477e-3 1,4430e-2]

Źródło: obliczenia własne.

pieniem do obliczeń, pierwotne dane pomnożono razy 100. Na podstawie szeregu $y_t = 100 \ln \text{CPI}_t$, dokonano estymacji parametrów w modelach: LL-GARCH (M_1), LL-GARCH-Student (M_2) i LL-SV (M_3), natomiast dla pierwszych różnic Δy_t użyto modele: RCA(2)-SV (RW) (M_4) i AR(2)-SV (M_5). W pierwszej linii w tabelach 3 i 4 znajdują się wartości oczekiwane, natomiast w drugiej linii umieszczono mediany rozkładów *a posteriori*. W nawiasach półokrągłych podano wartości odchyłeń standardowych, natomiast w nawiasach kwadratowych znajdują się 95% przedziały o najwyższej gęstości *a posteriori* (HPDI).

Otrzymane wyniki estymacji pozwalają stwierdzić, że rozważany szereg, zdecydowanie przemawia za występowaniem warunkowej heteroskedastyczności i to zarówno w równaniu obserwacji, jak i trendu. W modelach GARCH rozkłady *a posteriori* parametrów a_1 i b_1 są istotne statystycznie, ponieważ mają względnie małe rozproszenie, a ich przedziały HPD nie zawierają zera. Dodatkowo, suma punktowych ocen *a posteriori* parametrów a_1 i b_1 , w modelu LL-GARCH, jest bliska wartości jeden, co świadczy o silnej persystencji warunkowej wariancji. Z tego też powodu, w tabeli 3 zamieszczono również ich sumę wraz z poszczególnymi charakterystykami *a posteriori*. Jak łatwo zauważyć uzyskane punktowe oceny są bliskie jedynki, a górna granica przedziału HPD jest równa wartości jeden. Oznacza to, że jako dodatkowy model, w równaniu obserwacji i trendu można by było rozpatrzyć tzw. zintegrowany model GARCH, czyli IGARCH. W procesie IGARCH(1, 1) zachodzi $a_1 + b_1 = 1$, co oznacza, że jest to proces zintegrowany co do wariancji (Engle i Bollerslev,

1986). Bayesowską estymację i testowanie modeli IGARCH, dla polskich szeregów finansowych, przedstawiają Osiewalski i Pipień (1999).

W modelu LL-GARCH-Student można zauważyć, że rozkłady *a posteriori* parametru b_1 , w porównaniu z parametrami w modelu LL-GARCH, są zlokalizowane bardziej na lewo, co świadczy o mniejszej reakcji na opóźnione wartości warunkowej wariancji. Dotyczy to zwłaszcza parametru w równaniu obserwacji, gdzie wartość oczekiwana i mediana, w modelu LL-GARCH-Student, wynoszą odpowiednio 0,6090 i 0,6802, natomiast w modelu LL-GARCH są one równe 0,7431 i 0,7857. Oznacza to, że przyjęcie rozkładu o grubszych ogonach prowadzi do zmniejszenia persystencji zmienności. Dodatkowo, uwzględnienie w modelu GARCH warunkowych rozkładów *t*-Studenta, prowadzi do mniejszej precyzji. Pomimo pewnych różnic, w lokalizacji parametrów i rozproszeniu, oba modele GARCH dają podobny opis zmienności badanego szeregu. Także w modelach zmienności stochastycznej, średnia i mediana *a posteriori* parametru ρ przyjmują wartości zbliżone do jedynki, co wskazuje na występowanie bardzo silnej persystencji warunkowej wariancji, a lokalizacja rozkładu *a posteriori* jest podobna do tej przyjętej *a priori*. Trwałość zakłóceń losowych (szoków), w zmienności obserwowanego procesu mierzona za pomocą współczynnika $HL = \ln 0,5 / \ln \rho$ (por. Pajor, 2003), jest bardzo duża, rozproszona i cechuje się prawostronną asymetrią. Dolny kraniec przedziału obejmującego 80% całej masy *a posteriori* wynosi 5,73 miesiące, natomiast górny aż 138,59 miesięcy, z kolei mediana *a posteriori* jest równa 53,91 miesięcy. W przypadku szoków w zmienności trendu, rozpiętość przedziału jest jeszcze większa, ponieważ dolny i górny kraniec przedziału wynosi odpowiednio 10,98 i 227,08, natomiast mediana jest równa 95,39 miesięcy.

Mimo, że oceny punktowe współczynnika ρ są bliskie wartości jeden to masa rozkładu *a posteriori*, współczynnika autoregresji, w równaniu warunkowej wariancji, jest zlokalizowana w obszarze stacjonarności. Dzieje się tak, ponieważ charakteryzują się one małym rozproszeniem i niewielką rozpiętością przedziału HPD. Punktowe oceny parametru ρ w modelach SV są zgodne z wynikami, zamieszczonymi w artykule Grassiego i Proiettiego (2008), dla kwartalnych danych wskaźnika cen konsumenta CPI w Stanach Zjednoczonych.

Wartość oczekiwana *a posteriori* liczby stopni swobody, w modelu LL-GARCH-Student, w równaniu obserwacji kształtuje się na poziomie 2,4796, natomiast w równaniu trendu jest równa 2,4891. Dodatkowo, rozkłady te, są bardzo ciasno zlokalizowane na prawo od wartości $v = 2$, o czym świadczą zamieszczone w tabeli 3 przedziały HPD. Fakt, że niemal cała masa prawdopodobieństwa *a posteriori* jest skupiona poniżej wartości cztery oznacza, że dane silnie odrzucają hipotezę o istnieniu kurtozy rozkładów. Taka sytuacja wskazuje na rozkład warunkowy, wyraźnie różny od rozkładu normalnego oraz potwierdza występowanie bardzo ciężkich ogonów, co wiąże się z częstym występowaniem obserwacji nietypowych, w procesie obserwacji i trendu.

Badany szereg ma zagwarantowaną skończoną wariancję ponieważ, przyjęto rozkład *a priori*, którego gęstość przyjmuje niezerowe wartości na prawo od $v = 2$.

W tabeli 4 zamieszczono wyniki estymacji parametrów, w modelu autoregresyjnym, w którym współczynniki autoregresji podlegają procesowi błędzenia przypadkowego, natomiast wariancja opisana jest przez model stochastycznej zmienności. Zmienne w czasie współczynniki wykazują się małą zmiennością, ponieważ ich pierwsze przyrosty mają stosunkowo małą wariancję. Widać to poprzez fakt, że większość masy prawdopodobieństwa *a posteriori* wariancji dla ω_i^2 , dla $i = 0, 1, 2$, przypada na rejony położone w pobliżu zera, z jednocześnie względnie dużym rozproszeniem.

W tabeli 5 umieszczono brzegowe gęstości wektora obserwacji, obliczone przy użyciu metody Newtona-Rafferty'ego oraz otrzymany na tej podstawie ranking modeli. Uzyskane rezultaty informują, że w przypadku szeregu inflacji proponowany w artykule model lokalnego poziomu, w którym błędy w równaniu obserwacji i trendu opisane są za pomocą modelu GARCH (1,1) z warunkowym rozkładem *t*-Studenta, ma największą moc objaśniającą. Znacznie niższa, w porównaniu z innymi konkurencyjnymi modelami, wartość logarytmu brzegowej gęstości wektora obserwacji (-91,88) informuje, że dane wyraźnie wskazują na ten model, czyniąc pozostałe specyfikacje wysoce nieprawdopodobne. Drugim w kolejności, po modelu LL-GARCH-Student, jest model zaproponowany przez Stocka i Watsona, czyli model lokalnego poziomu z błędami typu SV. Model ze stałymi parametrami z wartością -141,72 ma dopiero czwarte miejsce w rankingu, po modelu LL-GARCH. Ostatnie miejsce w rankingu ma model autoregresyjny ze współczynnikami generowanymi przez proces błędzenia przypadkowego. Model ten zajmuje dopiero piąte miejsce, jedną pozycję niżej niż standardowy model AR.

Tabela 5

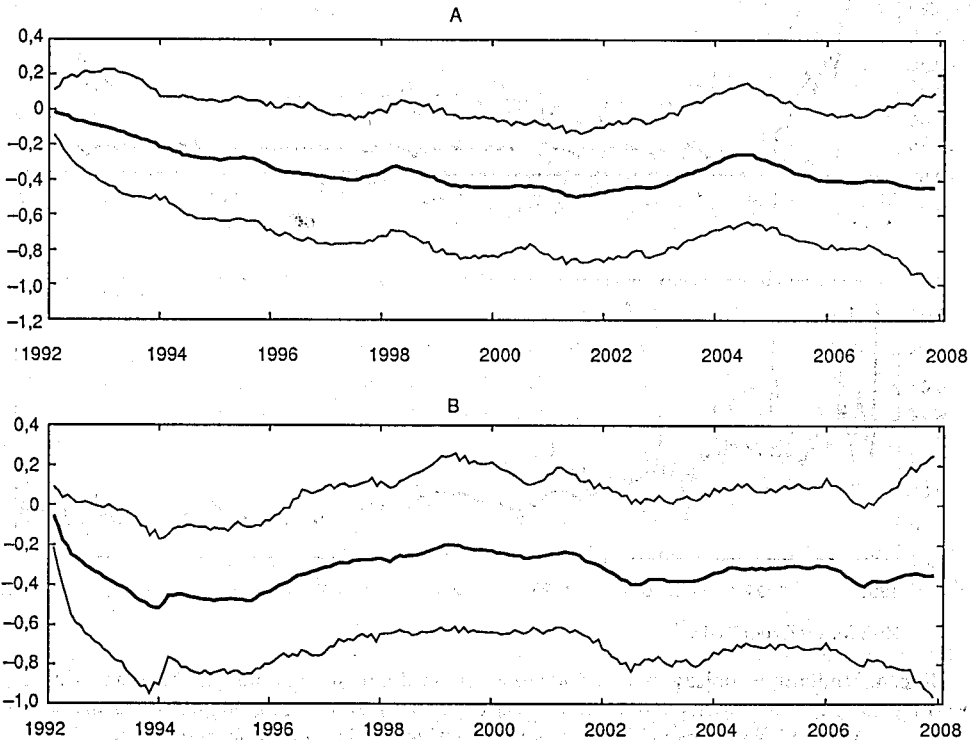
Logarytm brzegowej gęstości wektora obserwacji oraz ranking konkurujących modeli

	LL-GARCH	LL-GARCH-Student	LL-SV	RCA(2)-SV (RW)	AR(2)-SV
Logarytm brzegowej gęstości obserwacji	-134,14	-91,88	-126,13	-153,18	-141,72
Ranking modeli	3	1	2	5	4

Źródło: obliczenia własne.

Zbliżoną postać, do modelu RCA(2)-SV, miał model użyty przez Koopa i Pottera (2001). W ich badaniu uzyskał on znaczną przewagę nad konkurencyjnymi specyfikacjami. Okazuje się jednak, że w przypadku danych polskich, ma on najniższą pozycję w rankingu.

Stosunkowo małą zmienność współczynników autoregresji wydaje się potwierdzać rycina 3, na której zamieszczono wybrane charakterystyki rozkładów *a posteriori* pierwszego δ_{1t} (górny rysunek) i drugiego współczynnika autoregresji δ_{2t} dla $t = 1, \dots, 192$. Pogrubioną linią połączono mediany *a posteriori*, natomiast cieńszą linią zaznaczono dolną i górną granicę przedziału o najwyższej gęstości *a posteriori*. Poza początkowym okresem, ich rozpiętość jest stosunkowo stała. Mała rozpiętość przedziału HPD w pierwszym roku, jest najprawdopodobniej następstwem przyjęcia mało rozproszonych rozkładów *a priori* stanów początkowych. Punktowe oceny na poziomie wartości oczekiwanej współczynników autoregresji, w modelu ze stałymi parametrami, wyniosły odpowiednio $-0,3404$ i $-0,34931$, z błędami równymi $6,9283e-2$ i $6,3179e-2$. Są one zatem znacznie oddalone od zera i tym samym istotne statystycznie, co potwierdza występowanie zautokorelowanej struktury pierwszych przyrostów. Porównując je z ciągiem median, przedstawionym na rycinie 3, można zauważyć, że mediany te są zbliżone do punktowych ocen współczynników w standardowym modelu AR.

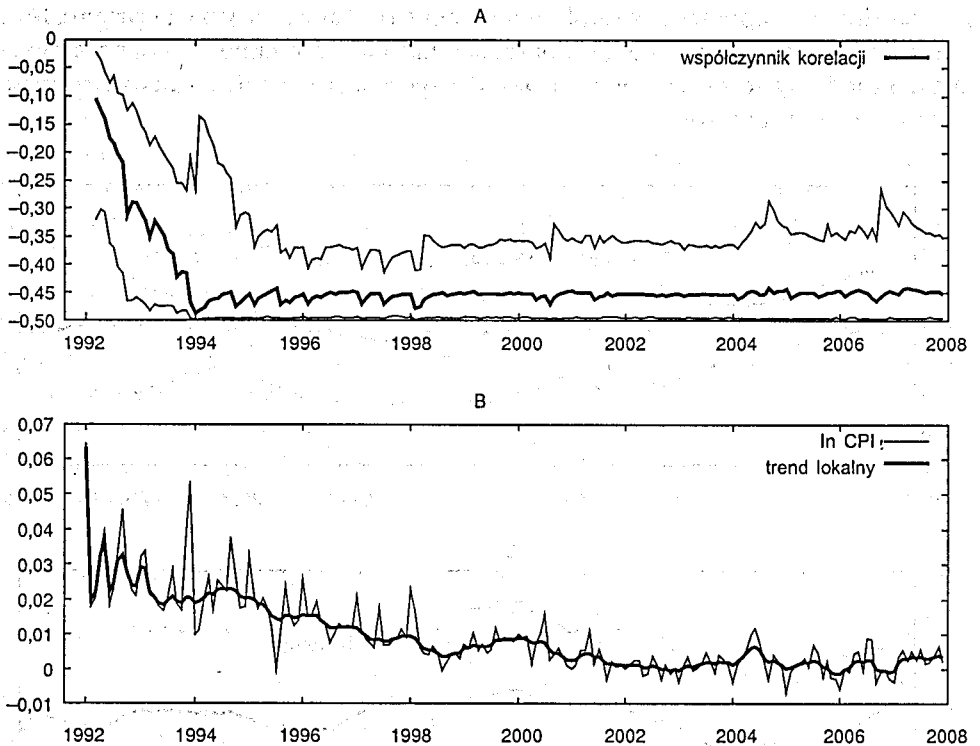


Źródło: obliczenia własne.

Ryc. 3. Pogrubioną linią oznaczono mediany współczynników autoregresji w modelu RCA-SV (RW): δ_{1t} — panel A i δ_{2t} — panel B oraz dolne i górne krańce 95% przedziałów o najwyższej gęstości *a posteriori*

Jak wspomniano w poprzedniej części artykułu, zmienność warunkowych wariacji ma ważne następstwa dla interpretacji współczynnika korelacji, który określa zależność między przyrostami inflacji a ich pierwszymi opóźnieniami. Jego wykres, obliczony za pomocą najbardziej prawdopodobnego modelu, czyli LL-GARCH-Student, przedstawia panel A na rycinie 4. Otrzymany ciąg współczynników jest zgodny z wynikami zaprezentowanymi w raporcie Cecchetti i in., 2007¹¹, ponieważ okresy wysokiej inflacji i niestabilności gospodarczej cechują się również znacznym wzrostem współczynnika korelacji.

Dla danych polskich, najbardziej interesujące wydają się lata 1992–1994, gdzie widać wyraźny spadek wartości współczynnika korelacji z $-0,1$ do $-0,5$, wraz z gwałtownym wzrostem rozpiętości przedziałów HPD w latach 1994–1995. Może to być spowodowane tym, że początek lat dziewięćdziesiątych



Źródło: obliczenia własne.

Ryc. 4. Mediany współczynnika autokorelacji (pogrubiona linia na panelu A) wraz z dolną i górną granicą 95% przedziałów o najwyższej gęstości *a posteriori* oraz logarytm wskaźnika CPI i trend lokalny (panel B). Wszystkie charakterystyki obliczono przy użyciu modelu LL-GARCH-Student

¹¹ Rycina 3.2 na s. 18.

był okresem dużych zmian w gospodarce, tj. radykalnej stabilizacji i liberalizacji makroekonomicznej, którym towarzyszyło zjawisko galopującej inflacji. W kolejnych latach, po roku 1994, współczynnik korelacji przyjmował wartości na znacznie niższym poziomie, równym około $-0,45$. Wskazuje to tym samym, poza początkowymi latami związanymi z okresem radykalnej transformacji, na stałą, niezmienną w czasie autoregresyjną strukturę inflacji.

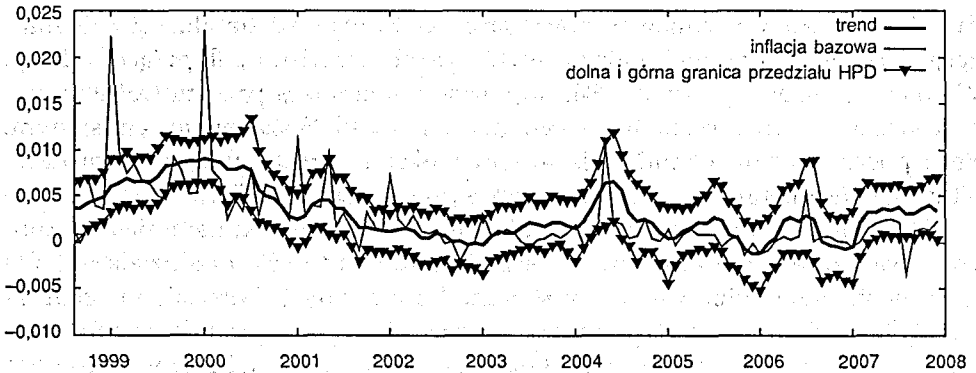
Punktowe oceny trendu prezentuje panel B na rycinie 4, natomiast warunkowe wariancje w równaniu obserwacji i trendu wraz z 95% przedziałami HPD są przedstawione odpowiednio na panelu A i B ryciny 6. Wszystkie te charakterystyki, podobnie jak wspomniany wcześniej współczynnik korelacji, estymowano za pomocą modelu LL-GARCH-Student. Jak widać, w badanym okresie, zarówno trend, jak i jego zmienność uległy znaczącej zmianie. Wartość trendu inflacji, obliczona na poziomie mediany *a posteriori*, zmalała z poziomu około 3% w roku 1992 do 0,3% w roku 2007, podczas gdy jego warunkowa wariancja z poziomu 4,96 w 1992 zmalała w tempie wykładniczym do blisko zera w roku 1994 i latach następnych.

Wydaje się interesujące porównanie ocen *a posteriori* nieobserwowalnego trendu, z inflacją bazową, której wartości podaje Narodowy Bank Polski. Jak wiadomo inflacji bazowa służy do oceny średnio- i długookresowego trendu ogólnego poziomu cen, ponieważ jej wartości ilustrują tendencje zmian oczyszczone z wahań okresowych oraz ze skutków szoków podaży, które najczęściej mają charakter przejściowy. Wiedza na temat inflacji bazowej umożliwia oszacowanie kierunku i skali wpływu prowadzonej przez bank centralny polityki pieniężnej¹².

Ponieważ bank centralny podaje wartości inflacji bazowej dopiero od stycznia 1998 roku, na rycinie 5 podano odpowiedni fragment ocen trendu oraz wartości logarytmów inflacji bazowej. Użyta inflacja bazowa to indeks cen po wyłączeniu cen żywności i energii, gdzie za podstawą wzięto poprzedni miesiąc. Punktowe wartości trendu zaznaczono pogrubioną linią łącząc mediany *a posteriori* zmiennej δ_t , natomiast cieńszą linią zaznaczono wartości inflacji bazowej. Dodatkowo na rycinie umieszczono również dolną i górną granicę 95% przedziałów o najwyższej gęstości *a posteriori* parametru δ_t .

Jak widać na powyższym rysunku uzyskany trend, wyznaczony za pomocą modelu LL-GARCH-Student, wskazuje na podobny kierunek zmian co wartości inflacji bazowej, ma on jednak w porównaniu z nią bardziej gładki przebieg, a jego przedziały ufności w zdecydowanej większości obejmują jej realizację. Potwierdza to zatem wiarygodność uzyskanych wyników badań, a zarazem daje możliwość ciekawej interpretacji modelu lokalnego poziomu,

¹² Narodowy Bank Polski — Internetowy Serwis Informacyjny, <http://www.nbp.pl/statystyka/bazowa/metodologia.pdf> (9.02.2009).

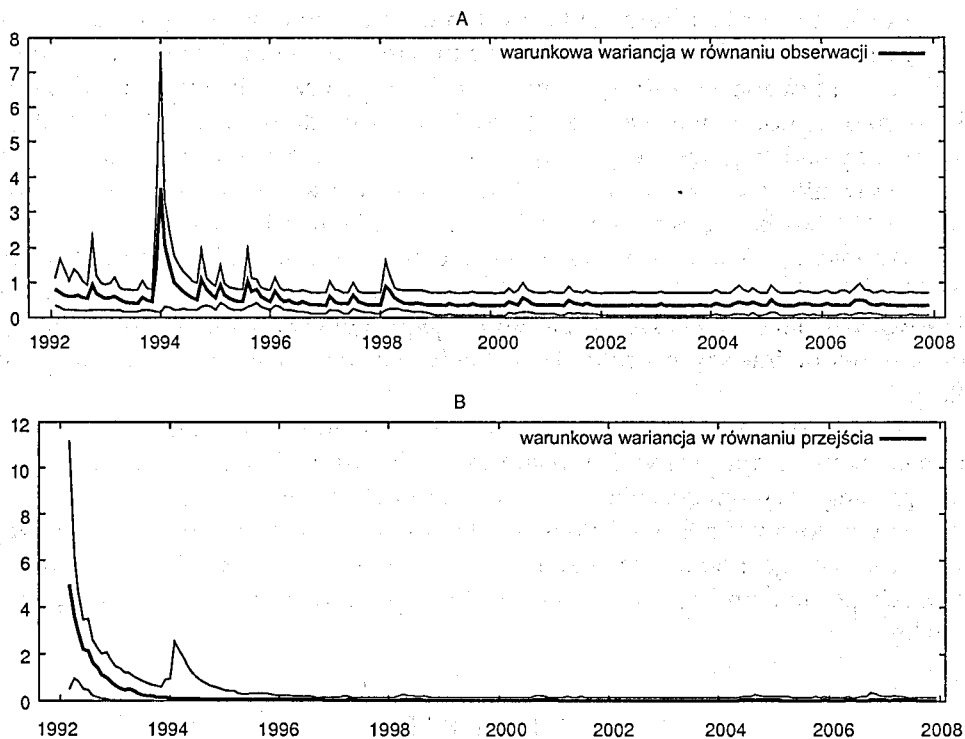


Źródło: obliczenia własne.

Ryc. 5. Punktowe oceny trendu wraz z dolną i górną granicą 95% przedziału HPD oraz wartości inflacji bazowej, gdzie za podstawę wzięto poziom cen z poprzedniego miesiąca

ponieważ zmienny parametr można interpretować jako długookresową tendencję inflacji.

Na panelu B ryciny 6 można zauważyć, że w latach 1992–1994 występował systematyczny spadek zmienności długookresowego trendu inflacji. Związane to było z rozpoczętym w 1990 roku i kontynuowanym w kolejnych latach programem stabilizacyjnym. Patrząc na panel A ryciny 6, można również dostrzec, że okres między rokiem 1992 a 1997 charakteryzuje się znacznie większą zmiennością niż lata późniejsze. W latach 1992–1993, polityka prowadzona przez bank centralny, kontrolowała względnie sprawne zasoby pieniężne i coraz sprawniej ograniczała inflację (Polański, 2005). Teza ta znajduje potwierdzenie na rycinie 6, ponieważ jak pokazuje wykres zmienności na panelu A, okres ten charakteryzował się względnie małą, w porównaniu z okresem późniejszym, zmiennością cen oraz wykładniczo malejącą zmiennością trendu (panel B). Największą zmienność obserwujemy pod koniec 1993 i na początku 1994 roku. Nagła zmiana poziomu cen mogła być skutkiem kilku niezależnie występujących, w tym okresie lub krótko przed nim, przyczyn. Po pierwsze, wystąpiło kilka jednorazowych egzogenicznych impulsów, takich jak „skokowa” dewaluacja złotego o 8% procent, wprowadzona w sierpniu 1993 (Polański, 1999) oraz „zamrożenie”, a następnie podniesienie stóp podatkowych w 1994 roku. Zdaniem ekonomistów (Bauc, 1995; Lutkowski, 1995; Grabia, 2003) skokowe dewaluacje złotego miały duże znaczenie dla kreacji pieniądza i doprowadziły do „wykolejenia” programu antyinflacyjnego. Po drugie, w drugiej połowie 1993 roku nastąpiła liberalizacja przepływów kapitałowych, która miała znaczny wpływ na utratę kontroli kształtowania podaży pieniądza przez bank centralny (Polański, 2005). Na problemy polityki stabilizacyjnej w tym okresie mógł się również wpłynąć fakt, że od 1994 polska gospodarka zaczęła być również



Źródło: obliczenia własne.

Ryc. 6. Mediany rozkładów a posteriori warunkowej wariancji w równaniu obserwacji (pogrubiona linia na panelu A) oraz w równaniu trendu (pogrubiona linia na panelu B), a także dolne i górne krańce 95% przedziałów o najwyższej gęstości a posteriori (zwykła linia). Wszystkie charakterystyki zostały obliczone przy użyciu modelu LL-GARCH-Student

w dużym stopniu zasilana ze źródeł zewnętrznych, pochodzących z nadwyżki w handlu przygranicznym (Polszakiewicz, 2005). Liberalizacja przepływów kapitałowych spowodowała, że polska gospodarka stanęła wobec tzw. dylematu niespójnego czworokąta. Polega on na tym, że władze monetarne nie mogą jednocześnie prowadzić autonomicznej polityki pieniężnej, stabilizować kursu walutowego, zapewnić pełnej mobilności przepływów kapitałowych oraz umożliwić pełną wymianę handlową (Polański, 2005; Polszakiewicz, 2005).

4. PODSUMOWANIE

Głównym celem tego artykułu było zastosowanie modeli, z parametrami generowanymi przez procesy stochastyczne, do opisu dynamiki inflacji w Polsce. Badania dotyczyły miesięcznych obserwacji wskaźnika cen konsumenta CPI w okresie od stycznia 1992 do grudnia 2007 roku.

Punktem wyjścia było wykorzystanie i porównanie mocy objaśniającej najnowszego modelu Stocka i Watsona z innymi modelami. Część z tych modeli była już używana wcześniej w badaniach empirycznych, inne zaś to nowe konstrukcje modelowe. Dotyczy to zwłaszcza modelu LL-GARCH-Student, który uzyskał największe prawdopodobieństwo *a posteriori*. Modelowanie procesów ukrytych o bardzo grubych ogonach jest zagadnieniem zupełnie nowym i okazało się ważne w wyjaśnianiu zmienności inflacji.

Współczynnik korelacji między przyrostami inflacji (CPI) a ich pierwszymi opóźnieniami dał możliwość wyodrębnienia dwóch zasadniczych podokresów. Pierwszy to lata 1992–1994, gdzie widać jego wyraźny spadek z $-0,1$ do $-0,5$ oraz okres późniejszy po roku 1994, kiedy nastąpiła stabilizacja struktury inflacji.

Na koniec należy również dodać, że poza dobrze udokumentowanymi przesłankami empirycznymi w najnowszej literaturze światowej, jak dotąd nie ma prostego wytłumaczenia losowości parametrów w modelach inflacji. Jako przyczyny losowości Stock i Watson (2007) podają szeroko rozumiane zmiany w strukturze gospodarki, założenia polityki monetarnej, rozwój rynków finansowych czy też zmianę natury potencjalnych szoków wpływających na gospodarkę.

BIBLIOGRAFIA

- Bauc J. 1995. *Inflacja polska w latach 1990–94 w świetle współczesnej teorii ekonomii*, Ekonomista 3, 441–461.
- Carlin B.P., Louis T.A. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, New York.
- Cecchetti S.G., Hooper P., Kasman B.C., Schoenholtz K.L., Watson M.W. 2007. *Understanding the evolving inflation process*, Report U.S. Monetary Policy Forum.
- Dickey D.A., Fuller W.A. 1979. *Distribution of the estimators for autoregressive time series with a unit root*, Journal of the American Statistical Association 74, 427–431.
- Doman M., Doman R. 2004. *Ekonometryczne modelowanie dynamiki polskiego rynku finansowego*, Wyd. AE w Poznaniu, Poznań 2004.
- Engle R.F., Bollerslev T. 1986. *Modelling the persistence of conditional variances*, Econometric Reviews 5, 1–50.
- Grabia T. 2003. *Polityka kursowa w Polsce w okresie transformacji jako instrument walki z inflacją*, Gospodarka w praktyce i teorii, Wyd. UE w Łodzi, Łódź, 1, 12, 48–58.
- Grassi S., Proietti T. 2008. *Has the Volatility of U.S. Inflation Changed and How?*, wersja niepublikowana.
- Koop G. 2003. *Bayesian Econometrics*, John Wiley & Sons.
- Koop G., Potter S. 2001. *Are apparent findings of nonlinearity due to structural instability in economic time series?*, The Econometrics Journal 4, 1, 37–55.
- Koop G., van Dijk H.K. 2000. *Testing for integration using evolving trend and seasonals models: A Bayesian approach*, Journal of Econometrics 97, 2, 261–291.
- Kotłowski J. 2006. *Pieniądz i ceny w gospodarce rynkowej. Analiza kointegracji sezonowej*, Wyd. SGH, Warszawa.

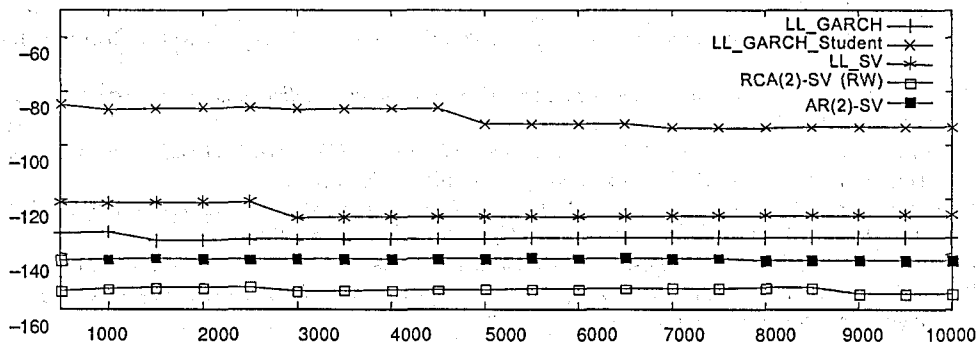
- Kwiatkowski J. 2009. *Bayesowska estymacja modelu lokalnego poziomu o rozkładach dopuszczających warunkowy rozkład t-Studenta i zmienną wariancję*, wersja niepublikowana.
- Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. 1992. *Testing the null hypothesis of stationarity against the alternative of a unit root*, *Journal of Econometrics* 54, 159–178.
- Lutkowski K. 1995. *Uwarunkowania polskiej inflacji i proces jej wygasania*, *Ekonomista* 3, 463–476.
- Majsterek M. 2008. *Wielowymiarowa analiza kointegracyjna w ekonomii*, Wyd. UŁ w Łodzi, Łódź.
- Osiewalski J. 2001. *Ekonometria bayesowska w zastosowaniach*, Wyd. AE w Krakowie, Kraków.
- Osiewalski J., Pipień M. 1999. *Bayesowskie testowanie modeli GARCH i IGARCH*, *Przegląd Statystyczny* 46, 1, 5–23.
- Osińska M. 2000. *Ekonometryczne modelowanie oczekiwań gospodarczych*, Wyd. UMK w Toruniu, Toruń.
- Pajor A. 2003. *Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych*, Wyd. AE w Krakowie, Kraków.
- Pellegrini S., Ruiz E., Espasa A. 2007. *The relationship between ARIMA-GARCH and unobserved component models with GARCH disturbances*, wersja niepublikowana.
- Pellegrini S., Ruiz E., Espasa A. 2008. *ARIMA-GARCH and unobserved component models with GARCH disturbances: Are their prediction intervals different?*, wersja niepublikowana
- Pipień M. 2006. *Wnioskowanie bayesowskie w ekonometrii finansowej*, Wyd. AE w Krakowie, Kraków.
- Polański Z. 1999. *Polityka kursu walutowego w Polsce w latach 90. Stabilizacja, konkurencyjność i przepływy kapitałowe*, *Ekonomista* 1–2, 135–154.
- Polański Z. 2005. *Przemiana i ciągłość: refleksje o polskiej polityce pieniężnej w okresie piętnastolecia transformacji ustrojowej*, *Gospodarka polska na przełomie wieków od A do Z*, red. Czarny E., NBP, Warszawa, 93–128.
- Polszakiewicz B. 2005. *Niespójny czworokąt rozważań systemowych w polskiej polityce stabilizacyjnej, Ład instytucjonalny w gospodarce*, red. Polszakiewicz, B., Boehlke, J., Wyd. UMK w Toruniu, Toruń, 3, 295–304.
- Przybylska-Kapuścińska W. 2006. *Ocena realizacji polityki pieniężnej Narodowego Banku Polskiego, Ład instytucjonalny w gospodarce*, red. Polszakiewicz, B., Boehlke, J., Wyd. UMK w Toruniu, Toruń, 2, 229–252.
- Stock J. H., Watson M.W. 2007. *Why has U.S. inflation become harder to forecast?* *Journal of Money, Credit, and Banking* 39, 3–33.
- Stock J. H., Watson M.W. 2008. *Phillips curve inflation forecasts*, wersja niepublikowana.
- Tsay R.S. 1987. *Conditional heteroscedastic time series models*, *Journal of the American Statistical Association* 82, 398.
- Welfe A. 1993. *Inflacja i rynek*, PWE, Warszawa.
- Welfe A. 2000. *Modelling inflation in Poland*, *Economic Modelling* 17, 375–385.
- Welfe A., Majsterek M. 2002. *Wage and prices inflation in Poland in the period of transition: The cointegration analysis*, *Economics of Planning* 35, 205–219.

DODATEK A

Analiza stabilności estymatora Newtona-Raftery'ego

W części 3, do obliczenia mocy objaśniającej konkurujących modeli oraz nadania im odpowiedniej rangi, posłużono się metodą Newtona-Raftery'ego. Jak wiadomo, główną wadą tego estymatora jest jego niestabilność, ponieważ nie spełnia on centralnego twierdzenia granicznego (Carlin i Louis, 2000). Z praktycznego punktu widzenia dzieje się tak, ponieważ bardzo małe wartości funkcji

wiarygodności w znaczny sposób wpływają na wielkość średniej harmonicznej, dlatego też ważne jest sprawdzenie jak kształtowały się, wraz ze wzrostem cykli, jego realizacje dla poszczególnych modeli. Rycina A.1 przedstawia przebieg brzegowej gęstości wektora obserwacji w zależności od liczby cykli. Jak widać zachowanie się estymatorów Newtona-Raftery'ego dla każdego modelu było stabilne.



Źródło: obliczenia własne.

Ryc. A.1. Wartości logarytmu estymatora Newtona-Raftery'ego wraz ze wzrostem liczby cykli

BIBLIOTEKA KA
im. Andrzeja Frycza Modrzowskiego

Dar/Wymiana WYDAWCA KA

Prenumerata

Data wpływu 12.05.2010



