POLSKA AKADEMIA NAUK
KOMITET STATYSTYKI I EKONOMETRII

PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

TOM 60

3

2013

WARSZAWA 2013

SPIS TREŚCI

POLSCY STATYSTYCY I MATEMATYCY

CONTENTS

POLISH STATISTICIANS AND MATHEMATICIANS

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

EMIL PANEK

„SŁABY" I „BARDZO SILNY" EFEKT MAGISTRALI W NIESTACJONARNEJ GOSPODARCE GALE'A Z GRANICZNĄ TECHNOLOGIĄ

J. von Neumann zaproponował liniowy model gospodarki (model von Neumanna, 1945-1946), w którym zdefiniował pojęcie stanu równowagi ekonomicznej oparte na idei równości technologicznych i ekonomicznych efektów produkcji. Dalsze badania nad modelem von Neumanna doprowadziły do powstania całej teorii neumannowskiej równowagi ekonomicznej. Badania w tym kierunku rozpoczęła między innymi grupa ekonomistów i matematyków pod kierunkiem Samuelsona (1960), który sformułował hipotezę, że w długich okresach czasu (horyzontach gospodarki) optymalne procesy wzrostu powinny przebiegać po ścieżkach wyznaczonych przez neumannowską równowagę ekonomiczną. Rozwój gospodarki wzdłuż takich ścieżek powinien być tym dłuższy, im dłuższy jest postulowany horyzont gospodarki. Obszar objęty neumannowską równowagą ekonomiczną przyjęło się nazywać magistralą (na wzór magistrali – autostrady – w ruchu drogowym).

Pionierska praca J. von Neumanna zapoczątkowała cały nurt badań, który w II połowie XX w. doprowadził do powstania w ekonomii matematycznej subdyscypliny dzisiaj znanej jako teoria magistral. W bogatej literaturze przedmiotu punkt ciężkości kładzie się na badanie tzw. efektu magistrali w różnych wariantach stacjonarnych gospodarek typu Neumanna – Gale'a – Leontiefa (ze stałą w czasie technologią; zob. np. bibliografię w pracy Panek, 2011). Natomiast nieliczne są prace poświęcone efektowi magistrali w niestacjonarnych gospodarkach ze zmienną technologią, zob. Gantz (1980), Joshi (1997), Keeler (1972). Wyzwanie takie podejmujemy w artykule. Inspiracją do jego napisania była publikacja autora (2013). Obowiązują oznaczenia stosowane w pracach Panek (2011, 2013).

1. MODEL NIESTACJONARNEJ GOSPODARKI GALE'A Z GRANICZNĄ TECHNOLOGIĄ. PODSTAWOWE ZAŁOŻENIA

Zakładamy, że czas t zmienia się skokowo, $t = 0,1,\ldots$. Zbiór $T = \{0,1,\ldots,t_1\}$ nazywamy horyzontem gospodarki, okres końcowy t_1 definiuje długość horyzontu T. W gospodarce Gale'a zużywa się i wytwarza n towarów, których liczba nie zmienia się w czasie. Przez $x(t) = (x_1(t),\ldots,x_n(t)) \geqq 0$ oznaczamy wektor towarów zużywanych (nazywamy go wektorem nakładów lub zużycia produkcyjnego), a przez

$y(t) = (y_1(t), \ldots, y_n(t)) \geqq 0$ wektor towarów wytwarzanych w gospodarce w okresie t (nazywamy go wektorem wyników lub produkcji). Jeżeli z wektora nakładów $x(t)$ możliwe jest wytworzenie wektora produkcji y(t), wtedy o parze $(x(t), y(t))$ mówimy, że opisuje technologicznie dopuszczalny proces produkcji (w okresie t). Zbiór $Z(t) \subset R_+^{2n}$ wszystkich technologicznie dopuszczalnych procesów produkcji tworzy tzw. przestrzeń produkcyjną w okresie t. Inkluzja $(x,y) \in Z(t)$ oznacza, że w świetle technologii jaką dysponuje gospodarka w okresie t, z wektora nakładów x możliwe jest wytworzenie wektora produkcji y. W razie potrzeby będziemy zamiennie stosować także zapis $(x(t), y(t)) \in Z(t)$. O przestrzeniach produkcyjnych $Z(t)$ zakładamy, że dla $t = 0,1,\ldots$ spełniają następujące warunki (zob. np. Panek, 2003, rozdz. 5):

(G1) $\forall (x,y) \in Z(t)\, \forall \lambda \geq 0 \left(\lambda(x,y) \in Z(t)\right)$
(warunek proporcjonalności nakładów i wyników),

(G2) $\forall (x^i, y^i) \in Z(t), i = 1,2 \ \left(x^1 + x^2, y^1 + y^2 \in Z(t)\right)$
(warunek addytywności procesów produkcyjnych),

(G3) $\forall (x,y) \in Z(t)\, (x = 0 \Rightarrow y = 0)$
(warunek „braku rogu obfitości),

(G4) $\forall (x,y) \in Z(t) \left(x' \geqq x \Rightarrow (x',y) \in Z(t)\right)$
(możliwość marnotrawstwa nakładów),

(G5) $\forall (x,y) \in Z(t)\, (0 \leqq y' \leqq y \Rightarrow (x,y') \in Z(t))$
(możliwość marnotrawstwa mocy produkcyjnych),

(G6) przestrzenie produkcyjne $Z(t)$ są zbiorami domkniętymi w R^{2n},

(G7) $Z(t) \subseteq Z(t+1) \subseteq Z$ i Z jest najmniejszym zbiorem domkniętym w R^{2n} zawierającym wszystkie przestrzenie produkcyjne $Z(t)$, spełniającym warunki **(G1)** – **(G6)** (rozwój technologii w gospodarce zwiększa z czasem jej możliwości wytwórcze).

Przestrzenie produkcyjne $Z(t)$, $t = 0,1,\ldots$, spełniające warunki **(G1)** – **(G6)** nazywamy gale'owskimi. Zbiór Z nazywamy graniczną przestrzenią produkcyjną (będziemy też zamiennie pisać, że przestrzeń Z opisuje graniczną technologię w niestacjonarnej gospodarce Gale'a). Zapis $(x,y) \in Z$ oznacza, że w świetle granicznej technologii, w niestacjonarnej gospodarce Gale'a, z nakładów x możliwe jest wytworzenie produkcji y. Gospodarkę, której przestrzeń produkcyjna $Z(t)$, Z spełniają warunki **(G1)** – **(G7)** nazywamy niestacjonarną gospodarką Gale'a z graniczną technologią.

□ **Twierdzenie 1**

Jeżeli przestrzenie produkcyjne $Z(t)$, Z spełniają warunki **(G1)** – **(G7)**, to

$$(x,y) \in Z \Leftrightarrow \exists\ (x(t), y(t)) \in Z(t)\,,\ t = 0,1,\ldots \left(\lim_t (x(t), y(t)) = (x,y)\right)$$

(tutaj i wszędzie dalej dla jednoznaczności przyjmujemy, że $\lim_t a(t) = a \Leftrightarrow \lim_t \|a(t) - a\| = 0$, $\|a\| = \sum_i |a_i|$).

Dowód. (⇒) Niech $(x,y) \in Z$). Utwórzmy ciąg $\{x(t), y(t)\}_{t=0}^{\infty}$, w którym $\forall t(x(t) = x)$ oraz

$$y(t) = \arg\min_{(x,y')\in Z(t)} \|y' - y\|. \tag{1}$$

Ciąg taki istnieje, gdyż dla każdego t zadanie (1) ma rozwiązanie (norma $\|\ \|$ jest bowiem funkcją ciągłą, a zbiory $\Omega_t(x) = \{y' | (x,y') \in Z(t)\}$ są zwarte). Pokażemy, że tak zbudowany ciąg jest zbieżny do granicy (x,y). Faktycznie, ponieważ $\forall t(x(t) = x)$, wystarczy wykazać, że $\lim_t y(t) = y$. Z inkluzji $Z(t) \subseteq Z(t+1) \subseteq ... \subseteq Z$ wnioskujemy, że

$$\|y(t+1) - y\| \le \|y(t) - y\|, \tag{2}$$

tj. ciąg $\{y(t)\}_{t=0}^{\infty}$ jest ograniczony, więc zawiera podciąg $\{y(t_k)\}_{k=1}^{\infty}$ zbieżny do pewnej granicy y^0. Załóżmy, że $y^0 \neq y$. Z definicji przestrzeni $Z(t), Z$ wynika, że

$$y^0 \le y \tag{3}$$

(przypominamy, że zapis $a \le b$ oznacza $a \leqq b$ oraz $a \neq b$). Niech $\tilde{y} = \frac{1}{2}(y^0 + y)$ Wówczas $\tilde{y} \le y$ i $(x,\tilde{y}) \in Z$ (zgodnie z **(G5)**). Jednocześnie

$$(x,\tilde{y}) \notin dom \bigcup_{t=0}^{\infty} Z(t)$$

(symbolem *dom A* oznaczamy najmniejszy zbiór domknięty zawierający A). Istotnie, gdyby $(x,\tilde{y}) \in dom \bigcup_{t=0}^{\infty} Z(t)$, to istniałby ciąg procesów $(x,\tilde{y}(t)) \in Z(t)$, $t = 0,1,\ldots$, zbieżny do $(x,\tilde{y})$, gdzie $\tilde{y} \ge y^0$. Zgodnie z (3)

$$\|\tilde{y} - y\| < \|y^0 - y\|,$$

co przeczy definicji (1) ciągu $\{y(t)\}_{t=0}^{\infty}$.

Zatem $\lim_k y(t_k) = y$. Wówczas na podstawie (2) wnioskujemy, że $\lim_t y(t) = y$.

(⇐) Ponieważ $(x(t), y(t)) \in Z(t) \subseteq Z$, $t = 0,1,\ldots$ oraz $\lim_t (x(t), y(t)) = (x,y)$ i graniczna przestrzeń produkcyjna Z jest domknięta w R^{2n}, zatem $(x,y) \in Z$.

■

Przestrzenie produkcyjne $Z(t), Z$ są stożkami domkniętymi w R_+^{2n} z wierzchołkami w 0. Interesują nas nietrywialnie procesy $(x,y) \neq 0$, tj. elementy zbiorów $Z(t)\backslash\{0\}, Z\backslash\{0\}$.

2. TECHNOLOGICZNA I EKONOMICZNA EFEKTYWNOŚĆ PRODUKCJI. RÓWNOWAGA VON NEUMANNA W GOSPODARCE Z GRANICZNĄ TECHNOLOGIĄ

Weźmy proces $(x,y) \neq 0$. Wówczas, zgodnie z **(G3)**, mamy $x \neq 0$. Liczbę

$$\alpha(x,y) = \max\{\alpha | \alpha x \leqq y\}$$

nazywamy wskaźnikiem technologicznej efektywności procesu (x,y). Funkcja α jest ciągła na $Z(t)\backslash\{0\}$ oraz $Z\backslash\{0\}$ i dodatnio jednorodna stopnia 0 (Panek, 2003, rozdz. 5, tw. 5.2.). Liczby

$$\alpha_{M,t} = \max_{(x,y)\in Z(t)\backslash\{0\}} \alpha(x,y)\,, \tag{4}$$

$$\alpha_M = \max_{(x,y)\in Z\backslash\{0\}} \alpha(x,y) \tag{5}$$

nazywamy odpowiednio:
- optymalnym wskaźnikiem technologicznej efektywności produkcji w niestacjonarnej gospodarce Gale'a w okresie t (z przestrzenią produkcyjną $Z(t)$),
- optymalnym wskaźnikiem technologicznej efektywności produkcji w niestacjonarnej gospodarce Gale'a z graniczną technologią (z graniczną przestrzenią produkcyjną Z).

□ **Twierdzenie 2**

Przy założeniach twierdzenia 1:

(i) zadania (4), (5) mają rozwiązania, tj.

$$\exists(\bar{x}(t),\bar{y}(t)) \in Z(t)\left(\alpha(\bar{x}(t),\bar{y}(t)) = \max_{(x,y)\in Z(t)\backslash\{0\}} \alpha(x,y) = \alpha_{M,t}\right),$$

$$\exists(\bar{x},\bar{y}) \in Z\left(\alpha(\bar{x},\bar{y}) = \max_{(x,y)\in Z\backslash\{0\}} \alpha(x,y) = \alpha_M\right),$$

(2i) $\forall t(\alpha_{M,t} \leq \alpha_{M,t+1} \leq \alpha_M)$.

Dowód. (i) Ponieważ funkcja α jest dodatnio jednorodna stopnia 0 (i ciągła) na $Z(t)\backslash\{0\}$, $Z\backslash\{0\}$, więc zadania (4), (5) są równoważne z zadaniami (odpowiednio)

$$\max_{(x,y)\in\Gamma(t)} \alpha(x,y) \tag{4'}$$

$$\max_{(x,y)\in\Gamma} \alpha(x,y) \tag{5'}$$

maksymalizacji ciągłej funkcji α na zwartych zbiorach

$$\Gamma(t) = \{(x,y) \in Z(t) \mid \|x,y\| = 1\},$$

$$\Gamma = \{(x,y) \in Z \mid \|x,y\| = 1\},$$

które mają rozwiązania (tw. Weierstrassa).

(2i) Zgodnie z **(G7)**

$$Z(t) \subseteq Z(t+1) \subseteq ... \subseteq Z,$$

stąd

$$\alpha_{M,t} = \max_{(x,y)\in Z(t)\backslash\{0\}} \alpha(x,y) = \alpha(\bar{x}(t),\bar{y}(t) \leq \max_{(x,y)\in Z(t+1)\backslash\{0\}} \alpha(x,y) = \alpha(\bar{x}(t+1),\bar{y}(t+1) \leq ...$$

$$\leq \max_{(x,y)\in Z\backslash\{0\}} \alpha(x,y) = \alpha(\bar{x},\bar{y}) = \alpha_M .$$

■

O parze wektorów $(\bar{x}(t),\bar{y}(t))$ mówimy, że tworzy optymalny proces produkcji w niestacjonarnej gospodarce Gale'a w okresie t; podobnie o parze wektorów $(\bar{x},\bar{y})$ mówimy, że tworzą optymalny proces produkcji w niestacjonarnej gospodarce Gale'a z graniczną technologią (graniczny optymalny proces produkcji w niestacjonarnej gospodarce Gale'a).

Ponieważ $\forall\lambda > 0$

$$\alpha(\bar{x}(t),\bar{y}(t)) = \alpha(\lambda\bar{x}(t),\lambda\bar{y}(t)) = \alpha_{M,t}$$

oraz

$$\alpha(\bar{x},\bar{y}) = \alpha(\lambda\bar{x},\lambda\bar{y}) = \alpha_M ,$$

zatem w niestacjonarnej gospodarce Gale'a optymalne procesy produkcji są określone z dokładnością do struktury.

O granicznej przestrzeni produkcyjnej Z zakładamy dodatkowo, że spełnia następujący warunek (tzw. warunek regularności):

(G8) $\exists(\bar{x},\bar{y}) \in Z\big(\alpha(\bar{x},\bar{y}) = \alpha_M \ \&\ \bar{y} > 0\big)$

(istnieje graniczny optymalny proces produkcji, w którym wytwarzane są wszystkie towary). Gospodarkę niestacjonarną Gale'a mającą tę własność nazywamy granicznie regularną. Z warunku **(G5)** wynika natychmiast, że w granicznie regularnej gospodarce Gale'a

$$\exists(\bar{x},\bar{y}) \in Z(\alpha_M \bar{x} = \bar{y}) \text{ oraz } \alpha_M > 0\,. \tag{6}$$

Mówiąc o granicznym optymalnym procesie produkcji mamy dalej na uwadze proces $(\bar{x},\bar{y})$ spełniający warunek (6).

O wektorze $\bar{s} = \dfrac{\bar{y}}{\|\bar{y}\|}$ mówimy, że charakteryzuje strukturę produkcji w granicznym optymalnym procesie. Z (6) wynika, że wektor ten charakteryzuje także strukturę nakładów w takim procesie:

$$\frac{\bar{x}}{\|\bar{x}\|} = \frac{\alpha_M^{-1}\bar{y}}{\|\alpha_M^{-1}\bar{y}\|} = \frac{\alpha_M^{-1}\bar{y}}{\alpha_M^{-1}\|\bar{y}\|} = \frac{\bar{y}}{\|\bar{y}\|} = \bar{s}\,.$$

Półprostą

$$N = \{\lambda\,\bar{s} \mid \lambda > 0\} \tag{7}$$

nazywamy magistralą produkcyjną (promieniem von Neumanna) w niestacjonarnej gospodarce Gale'a z graniczną technologią.

Niech $p = (p_1,\ldots,p_n) \geq 0$ będzie wektorem cen towarów oraz (x,y) dowolnym procesem produkcji. Liczbę

$$\beta(x,y,p) = \frac{\langle p,y\rangle}{\langle p,x\rangle}$$

(tam gdzie jest określana) nazywamy wskaźnikiem ekonomicznej efektywności procesu (x,y) przy cenach p (tutaj i wszędzie dalej $\langle a,b\rangle$ oznacza iloczyn skalarny wektorów a,b : $\langle a,b\rangle = \sum_i a_i b_i$).

Jeżeli gospodarka Gale'a jest granicznie regularna, to

$$\exists\bar{p} \geq 0 \quad \forall(x,y) \in Z\big(\langle \bar{p},y\rangle - \alpha_M \langle \bar{p},x\rangle \leq 0\big) \tag{8}$$

oraz

$$\beta(\bar{x},\bar{y},\bar{p}) = \max_{(x,y)\in Z\setminus\{0\}} \beta(x,y,\bar{p}) = \alpha(\bar{x},\bar{y}) = \alpha_M \tag{9}$$

(zob. Panek, 2003, tw. 5.3). Wektor $\bar{p}$ nazywamy wektorem cen von Neumanna w niestacjonarnej gospodarce Gale'a z graniczną technologią. O trójce $\{\alpha_M, (\bar{x},\bar{y}), \bar{p}\}$ mówimy, że charakteryzuje niestacjonarną gospodarkę Gale'a z graniczną technologią w równowadze von Neumanna. Dochodzi w niej do zrównania ekonomicznej efektywności produkcji z efektywnością technologiczną na najwyższym poziomie jaki może osiągnąć niestacjonarna gospodarka Gale'a z graniczną technologią.

W celu zapewnienia jednoznaczności magistrali produkcyjnej (7) zakładamy, że

(G9) $\forall (x,y)\in Z\setminus\{0\}\ \left(x\notin N \Rightarrow \beta(x,y,\bar{p}) = \dfrac{\langle \bar{p},y\rangle}{\langle \bar{p},x\rangle} < \alpha_M\right)$

(efektywność ekonomiczna jakiegokolwiek procesu poza magistralą jest niższa od optymalnej). Warunek $x \notin N$ jest równoważny z $\dfrac{x}{\|x\|} \neq \bar{s}$.

□ **Twierdzenie 3.** (Radner, 1961)

Jeżeli zachodzą warunki **(G1) – (G9)**, to

$$\forall \varepsilon > 0 \quad \exists \delta_\varepsilon > 0 \quad \forall (x,y)\in Z\left(\left\|\frac{x}{\|x\|} - \bar{s}\right\| \geq \varepsilon \Rightarrow \beta(x,y,\bar{p}) = \frac{\langle \bar{p},y\rangle}{\langle \bar{p},x\rangle} \leq \alpha_M - \delta_\varepsilon\right)$$

■

Twierdzenie to gra istotną rolę przy dowodzi magistralnych własności optymalnych procesów wzrostu, o których mowa w kolejnych twierdzeniach 4, 5.

3. WZROST

Odwzorowanie (multifunkcję) $a_t : R_+^n \to 2^{R_+^n}$ postaci

$$a_t(x) = \{y \mid (x,y)\in Z(t)\}$$

nazywamy przekształceniem technologicznym w niestacjonarnej gospodarce Gale'a w okresie *t* (generowanym przez przestrzeń produkcyjną *Z*(*t*)). Podobnie definiujemy przekształcenie technologiczne $a : R_+^n \to 2^{R_+^n}$,

$$a(x) = \{y \mid (x,y)\in Z\}$$

(generowane przez graniczną przestrzeń produkcyjną *Z*).

Jeżeli $Z(t)$ jest gale'owską przestrzenią produkcyjną (spełniającą warunki **(G1)** – **(G6)**), to przekształcenie technologiczne a_t ma następujące własności (Panek, 2003, tw.5.1):

- $\forall x \geqq 0 \quad \forall \lambda > 0 \left(a_t(\lambda x) = \lambda a(x)\right)$ (jest dodatnio jednorodne stopnia 1),
- $\forall x^1, x^2 \geqq 0 \quad \left(a_t(x^1 + x^2) \subseteq a_t(x^1) + a(x^2)\right)$ (jest póładdytywne),
- $a_t(0) = \{0\}$,
- $\forall (x, y) \in Z(t) \quad (y \in a_t(x) \,\&\, 0 \leqq y' \leqq y \Rightarrow y' \in a_t(x))$ (możliwość marnotrawstwa mocy produkcyjnych),
- $\forall (x, y) \in Z(t) \qquad (y \in a_t(x) \,\&\, x' \geqq x \Rightarrow y \in a_t(x'))$ (możliwość marnotrawstwa nakładów),
- $\forall x \geqq 0$ zbiory (obrazy) $a_t(x)$ są wypukłe i zwarte,
- przekształcenie a_t jest półciągłe (z góry), tzn. spełnia warunek

$$\forall (x^i, y^i)_{i=1}^{\infty} \left(y^i \in a_t(x^i) \,\&\, (x^i, y^i) \underset{i}{\rightarrow} (\bar{x}, \bar{y}) \Rightarrow \bar{y} \in a_t(\bar{x}) \right).$$

Podobne własności ma przekształcenie technologiczne a generowane przez graniczną przestrzeń produkcyjną Z spełniającą warunki **(G1)-(G7).**
Ponadto łatwo pokazać, że

- $\forall t \;\forall x \geqq 0 \;\; (a_t(x) \subseteq a_{t+1}(x) \subseteq a(x))$.

Weźmy dowolny ciąg procesów $(x(t), y(t)) \in Z(t)$, $t \in T = \{0, 1, \ldots t_1\}$. Standardowo zakładamy, że w horyzoncie T nakłady w okresie następnym mogą pochodzić tylko z produkcji wytworzonej w okresie poprzednim[1]: $x(t+1) \leqq y(t), \; t = 0,1, \ldots t_1 - 1$, co przy warunkach **(G5)**, **(G6)** prowadzi do inkluzji $(y(t), y(t+1)) \in Z(t+1), \; t = 0,1,\ldots t_1 - 1$, lub równoważnie:

$$y(t+1) \subseteq a_{t+1}(y(t)), \; t = 0,1,\ldots t_1 - 1. \tag{10}$$

Ustalmy wektor produkcji y^0 wytworzonej w gospodarce Gale'a w okresie początkowym $t = 0$:

$$y(0) = y^0 \geq 0. \tag{11}$$

O ciągu wektorów produkcji $\{y(t)\}_{t=0}^{t_1}$ spełniającym warunki (10) – (11) mówimy, że opisuje (y^0, t_1) – dopuszczalny proces wzrostu w niestacjonarnej gospodarce Gale'a. Przyjmijmy oznaczenia:

---

[1] Jest to wariant modelu Gale'a gospodarki zamkniętej. W rozbudowanych wersjach modelu warunek ten można osłabić.

$$R_{y^0,0} = \{y^0\},\ R_{y^0,t} = \bigcup_{x \in R_{y^0,t-1}} a_t(x),\quad t = 1,2,...t_1$$

($R_{y^0,t}$ jest zbiorem wszystkich wektorów produkcji możliwych do wytworzenia w okresie t przez niestacjonarną gospodarkę Gale'a z początkową produkcją $y(0) = y^0$).

Jeżeli spełnione są warunki **(G1)** – **(G6)**, to $\forall t_1 < +\infty\ \forall y^0 \geq 0$ zbiory $R_{y^0,t}$ są niepuste, zwarte i wypukłe. Innymi słowy, istnieją (y^0, t_1) dopuszczalne procesy wzrostu (Panek, 2003, lemat 5.1).

4. EFEKT MAGISTRALI

Rozpatrzmy następujące zadanie maksymalizacji wartości produkcji, mierzonej w cenach von Neumanna, wytworzonej w niestacjonarnej gospodarce Gale'a w ostatnim okresie horyzontu $T = \{0,1,\ldots t_1\}$:

$$\begin{aligned} &\max \langle \overline{p}, y(t_1) \rangle \\ &\text{p.w.(10) – (11)} \\ &\text{(wektor } y^0 \text{ ustalony).} \end{aligned} \tag{12}$$

Zadanie to ma rozwiązanie $\{y^*(t)\}_{t=0}^{t_1}$, które będziemy nazywać $(y^0, t_1, \overline{p})$ – optymalnym procesem wzrostu w niestacjonarnej gospodarce Gale'a[2]. Przez $s^*(t) = \dfrac{y^*(t)}{\|y^*(t)\|}$ oznaczamy strukturę produkcji w $(y^0, t_1, \overline{p})$ – optymalnym procesie wzrostu w okresie t.

W poniższym twierdzeniu, mówiącym o „magistralnej" stabilności procesów optymalnych, istotną rolę gra możliwość dojścia gospodarki do magistrali:

(G10) Istnieje taki $(y^0, \breve{t})$ – dopuszczalny proces wzrostu $\{\widetilde{y}(t)\}_{t=0}^{\breve{t}}$, gdzie $\breve{t} < t_1$, że

$$\alpha(\widetilde{y}(\breve{t}-1), \widetilde{y}(\breve{t})) = \alpha_M . \tag{13}$$

---

[2] Zadanie (12) ma rozwiązanie wtedy i tylko wtedy, gdy istnieje rozwiązanie zadania

$$\max \langle \overline{p}, y \rangle$$
$$y \in R_{y^0,t_1}$$

(maksymalizacji funkcji liniowej, a więc ciągłej, na zwartym zbiorze R_{y^0,t_1}). Zadanie to ma oczywiście rozwiązanie, zatem istnieje również rozwiązanie zadania (12).

□ **Twierdzenie 4**[3] („Słabe" twierdzenie o magistrali)

Jeżeli niestacjonarna gospodarka Gale'a spełnia warunki **(G1)-(G10)**, to dla dowolnej liczby $\varepsilon > 0$ istnieje taka liczba naturalna k_ε, że liczba okresów, w których struktura produkcji w $(y^0, t_1, \overline{p})$ – optymalnym procesie spełnia warunek

$$\left\| s^*(t) - \overline{s} \right\| \geq \varepsilon \tag{14}$$

nie przekracza k_ε. Liczba k_ε nie zależy od długości horyzontu T.

Dowód. Weźmy $(y^0, t_1, \overline{p})$ optymalny proces $\{\widetilde{y}(t)\}_{t=0}^{\breve{t}}$ (rozwiązanie zadania (12)). Ponieważ dla każdego t warunek $y^*(t+1) \in (y^*(t))$ jest równoważny z

$$(y^*(t),\ y^*(t+1)) \in Z(t+1) \subseteq Z,$$

więc zgodnie z (8)

$$\langle \overline{p}, y^*(t+1) \rangle \leq \alpha_M \langle \overline{p}, y^*(t) \rangle,\ t = 0,1,...t_1 - 1. \tag{15}$$

Oznaczmy przez L_ε zbiór tych okresów $t \in T$, w których zachodzi nierówność (14). Niech l_ε będzie liczbą elementów zbioru L_ε. Wówczas, zgodnie z twierdzeniem 3,

$$\langle \overline{p}, y^*(t+1) \rangle \leq (\alpha_M - \delta_\varepsilon) \langle \overline{p}, y^*(t) \rangle,\ \text{ dla } t \in L_\varepsilon. \tag{16}$$

Z (15), (16) otrzymujemy:

$$\langle \overline{p}, y(t_1) \rangle \leq \alpha_M^{t_1 - l_\varepsilon}\ (\alpha_M - \delta_\varepsilon)^{l_\varepsilon} \langle \overline{p}, y^0 \rangle. \tag{17}$$

Zgodnie z warunkiem **(G10)** istnieje taki $(y^0, \breve{t})$ – dopuszczalny proces wzrostu $\{\widetilde{y}(t)\}_{t=0}^{\breve{t}}$, $\breve{t} < t_1$, że $\alpha(\widetilde{y}(\breve{t}-1), \widetilde{y}(\breve{t})) = \alpha_M$. Ponieważ $\widetilde{y}(\breve{t}) \in a_{\breve{t}}(\widetilde{y}(\breve{t}-1))$, więc $(\widetilde{y}(\breve{t}-1), \widetilde{y}(\breve{t})) \in Z(\breve{t})$. Z (13) oraz z warunku (G5) wynika wówczas, że $(\widetilde{y}(\breve{t}-1), \alpha_M \widetilde{y}(\breve{t}-1)) \in Z(\breve{t})$, czyli $(\widetilde{y}(\breve{t}-1) \in N$, tzn. $\widetilde{y}(\breve{t}-1) = \sigma\overline{s}$ dla pewnej liczby $\sigma > 0$. Stąd, zważywszy na **(G7)**, otrzymujemy (y^0, t_1) – dopuszczalny proces

$$y(t) = \begin{cases} \widetilde{y}(t), & dla\ \ t = 0,1,...,\breve{t}-1 \\ \sigma\overline{s}\ \alpha_M^{t-\breve{t}+1}, & dla\ t = \breve{t}, \breve{t}+1,...,t_1 \end{cases}$$

[3] Jest to dostosowana do specyfiki niestacjonarnej gospodarki Gale'a z graniczną technologią wersja twierdzenia 5.8 udowodnionego w pracy Panek (2003).

oraz nierówność:

$$\langle \overline{p}, y^*(t_1)\rangle \geq \langle \overline{p}, y(t_1)\rangle = \sigma\alpha_M^{t_1-\breve{t}+1}\langle \overline{p}, \overline{s}\rangle > 0. \tag{18}$$

Łącząc warunki (17), (18) dochodzimy do nierówności

$$\alpha_M^{t_1-l_\varepsilon}\,(\alpha_M-\delta_\varepsilon)^{l_\varepsilon}\langle \overline{p}, y^0\rangle \geq \sigma\alpha_M^{t_1-\breve{t}+1}\langle \overline{p}, \overline{s}\rangle > 0,$$

z której otrzymujemy oszacowanie górnego ograniczenia liczby l_ε:

$$l_\varepsilon \leq \frac{\ln A}{\ln\alpha_M - \ln(\alpha_M-\sigma_\varepsilon)} = B,$$

gdzie $A = \dfrac{\alpha_M^{\breve{t}-1}\langle \overline{p}, y^0\rangle}{\sigma\langle \overline{p}, \overline{s}\rangle}$. W charakterze liczby k_ε wystarczy wziąć najmniejszą liczbę naturalną większą od $\min\{0,B\}$. Liczba ta nie zależy od t_1 (tj. od długości horyzontu T).

■

W myśl twierdzenia, optymalne procesy wzrostu w niestacjonarnej gospodarce Gale'a z graniczną technologią „prawie zawsze" (poza pewną skończoną liczbą okresów, niezależną od długości horyzontu, w którym badamy funkcjonowanie gospodarki) przebiegają dowolnie blisko magistrali wyznaczanej przez graniczną technologię. Na magistrali gospodarka osiąga najwyższe tempo wzrostu.

Prostą (aczkolwiek nietrywialną) konsekwencją twierdzenia 4 jest poniższa wersja „bardzo silnego" twierdzenia o magistrali z graniczną technologią.

□ **Twierdzenie 5** („Bardzo silne" twierdzenie o magistrali)

Jeżeli w niestacjonarnej gospodarce Gale'a spełniającej warunki **(G1)-(G9)** $(y^0, t_1, \overline{p})$ – optymalny proces wzrostu w pewnym okresie $\breve{t} < t_1$ spełnia warunek $\alpha(y^*(\breve{t}), y^*(\breve{t}+1)) = \alpha_M$ (prowadzi do magistrali N), to

$$\forall\, t \in \{\breve{t}+1, \ldots, t_1-1\}\ (y^*(t) \in N).$$

Dowód. Optymalny proces $\{y^*(t)\}_{t=0}^{t_1}$ z założenia w okresie $\breve{t} < t_1$ prowadzi do magistrali. Istnieje więc taka liczba $\sigma > 0$, że $(y^*(\breve{t}) = 65$ oraz (y^0, t_1) – dopuszczalny proces wzrostu $\{\widetilde{y}(t)\}_{t=0}^{t_1}$ postaci:

$$\widetilde{y}(t) = \begin{cases} y^*(t), & dla \quad t = 0,1,\ldots,\breve{t}-1 \\ \sigma\overline{s}\ \alpha_M^{t-\breve{t}}, & dla \quad t = \breve{t}, \breve{t}+1,\ldots,t_1 \end{cases}$$

(dowód przebiega analogicznie jak w twierdzeniu 4).

Wówczas

$$\langle \overline{p}, y^*(t_1)\rangle \geq \langle \overline{p}, \widetilde{y}(t_1)\rangle = \sigma\alpha_M^{t_1-\breve{t}}\langle \overline{p}, \overline{s}\rangle > 0. \tag{19}$$

Z drugiej strony, postępując jak przy dowodzie twierdzenia 4, z warunku (15) otrzymujemy:

$$\langle \overline{p}, y^*(t_1)\rangle \leq \alpha_M\langle \overline{p}, y^*(t_1-1)\rangle \leq \ldots \leq \alpha_M^{t_1-\breve{t}}\langle \overline{p}, y^*(\breve{t})\rangle = \sigma\alpha_M^{t_1-\breve{t}}\langle \overline{p}, \overline{s}\rangle. \tag{20}$$

Załóżmy, że w pewnym okresie $\tau \in \{\breve{t}+1, \ldots, t_1-1\}$:

$$y^*(\tau) \notin N, \text{ tzn. } \exists\varepsilon > 0 \left(\left\| \frac{y^*(\tau)}{\left\| y^*(\tau)\right\|} - \overline{s} \right\| \geq \varepsilon \right).$$

Wówczas, zgodnie z twierdzeniem 3

$$\exists\delta_\varepsilon > 0 \ (\langle \overline{p}, y^*(\tau+1)\rangle \leq (\alpha_M - \delta_\varepsilon)\langle \overline{p}, y^*(\tau)\rangle),$$

co łącznie z (20) prowadzi do nierówności:

$$\langle \overline{p}, y^*(t_1)\rangle \leq \sigma\alpha_M^{t_1-\breve{t}-1}(\alpha_M - \delta_\varepsilon)\langle \overline{p}, \overline{s}\rangle. \tag{21}$$

Z (19), (21) otrzymujemy warunek

$$\sigma\alpha_M^{t_1-\breve{t}-1}(\alpha_M - \delta_\varepsilon)\langle \overline{p}, \overline{s}\rangle \geq \sigma\alpha_M^{t_1-\breve{t}}\langle \overline{p}, \overline{s}\rangle > 0,$$

z którego wynika, że $\delta_\varepsilon \leq 0$. Otrzymana sprzeczność zamyka dowód.

■

Twierdzenie 5 głosi, że „wejście" optymalnego procesu wzrostu w pewnym okresie na magistralę w niestacjonarnej gospodarce Gale'a z graniczną technologią skutkuje jej pobytem na magistrali wszędzie dalej, za wyjątkiem ewentualnie ostatniego okresu horyzontu *T*.

Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

[1] Gantz D. T., (1980), A Strong Turnpike Theorem for a Nonstationary von Neumann-Gale Production Model, *Econometrica*, 48 (7), 1777-90.
[2] Joshi S., (1997), Turnpike Theorems in Nonconvex Nonstationary Envirenments, *Int. Econ. Rev.*, 38 (1), 225-248.
[3] Keeler E. B., (1972), A Twisted Turnpike, *Int. Econ. Rev.*, 13 (1), 160-166.
[4] Panek E., (2003), *Ekonomia matematyczna*, Wydawnictwo AE w Poznaniu.
[5] Panek E., (2011), O pewnej prostej wersji „słabego" twierdzenia o magistrali w modelu von Neumanna, *Przegląd Statystyczny*, 58 (1-2), 75-87.
[6] Panek E., (2013), Niestacjonarny model von Neumanna z graniczną technologią*, Studia Oeconomica Posnaniensia*, 1 (1), 49-68.
[7] Radner R., (1961), Paths of Economic Growth that are Optimal with Regard only to Final States: A Turnpike Theorem, *Rev. Econ. Studies*, 28 (2), 98-104.
[8] Samuelson P. A., (1959), Efficient Path of Capital Accumulation in Terms of the Calculus of Variations, w: Arrow K. J., Karlin S., Suppes P., (red.), *Mathematical Methods in the Social Sciences*, Stanford, California, 77-88.
[9] Von Neumann J., (1945-1946), A Model of General Economic Equilibrium, *Rev. Econ. Studies*, 13 (1), 1-9.

„SŁABY" I „BARDZO SILNY" EFEKT MAGISTRALI W NIESTACJONARNEJ GOSPODARCE GALE'A Z GRANICZNĄ TECHNOLOGIĄ

S t r e s z c z e n i e

W bogatej literaturze z teorii magistral zdecydowana większość prac poświęcona jest magistralnym własnościom optymalnych procesów wzrostu w stacjonarnych gospodarkach typu Neumanna – Gale'a – Leontiefa.

Do bardzo nielicznych należą publikacje, których autorzy podejmują próbę zbadania własności optymalnych procesów wzrostu w gospodarkach niestacjonarnych (ze zmienną technologią). Zadanie to podjęto w artykule na gruncie niestacjonarnej gospodarki Gale'a z technologią graniczną.

Słowa kluczowe: niestacjonarny model Gale'a, technologia graniczna, równowaga von Neumanna, magistrala produkcyjna

„WEAK" AND „VERY STRONG" TURNPIKE EFFECT IN THE NON-STATIONARY GALE ECONOMY WITH LIMIT TECHNOLOGY

A b s t r a c t

In the rich literature on the turnpike theory the vast majority of papers consider turnpike properties of optimal growth processes in the stationary Neumann-Gale-Leontief economies.

There are only few papers, in which their authors attempt to explore characteristics of the optimal growth processes in the non-stationary economies (with changeable technology). This objective has been undertook in this paper with the non-stationary Gale economy with the limit technology.

Keywords: non-stationary Gale model, limit technology, von Neumann equilibrium, production turnpike

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

ANDRZEJ SOKOŁOWSKI, SABINA DENKOWSKA,
KAMIL FIJOREK, MARCIN SALAMAGA

ANALIZA MOCY WYBRANYCH TESTÓW JEDNORODNOŚCI CZASÓW TRWANIA DLA POPULACJI O ROZKŁADZIE WEIBULLA[1]

1. WPROWADZENIE

Badanie niektórych zjawisk medycznych, demograficznych, społecznych, gospodarczych czy politycznych wymaga przeprowadzenia tzw. *analizy czasu trwania*. Termin ten w zależności od obszaru zastosowania ma również inne określenia: analiza przeżycia, dożycia (np. w medycynie, demografii, biologii), analiza przejścia (w ekonomii, naukach społecznych), analiza niezawodności (w naukach technicznych) (por. np. Balicki, 2006). Początkowo metody analizy trwania zjawisk znajdowały zastosowanie w statystyce medycznej i demografii, ale z czasem weszły do kanonu narzędzi badawczych innych dyscyplin naukowych. W szczególności w ostatnim czasie obserwuje się wzrost zainteresowania tymi metodami w badaniu zjawisk społeczno-gospodarczych, np. w zagadnieniach dotyczących wchodzenia przedsiębiorstw na rynek, czasu istnienia firm, rynku pracy, rynku nieruchomości, itd.

Dane wykorzystywane w analizie trwania mogą mieć charakter kompletnych bądź niekompletnych danych. Jedną z przyczyn niekompletności danych jest zjawisko cenzurowania jednostek wiążące się z ich eliminacją z pola obserwacji. Przykładowo podczas śledzenia zbioru obiektów mogą pojawić się jednostki, którym będzie towarzyszyć zdarzenie kończące ich obserwację lub też jednostki, w przypadku których takie zdarzenie nie wystąpi do zakończenia procesu obserwacji. W tej drugiej sytuacji jednostki nazywamy cenzurowanymi. Cenzurowanie może się odbywać ze względu na czas (ang. *time censoring*) lub ze względu na zakończenie badania w określonym terminie. Takie cenzurowanie nazywamy cenzurowaniem typu I (por. Balicki, 2006).

Wyróżnia się trzy podstawowe typy cenzurowania danych ze względu na czas (cenzurowanie I typu): cenzurowanie prawostronne, lewostronne oraz obustronne (przedziałowe). W praktyce najczęściej spotykane jest cenzurowanie prawostronne. Może mieć ono miejsce w jednej z dwóch sytuacji (por. Deszczyńska, 2011):

[1] Artykuł zawiera wybrane wyniki uzyskane w ramach Badań Statutowych prowadzonych w Katedrze Statystyki Uniwersytetu Ekonomicznego w Krakowie w 2012 r. (umowa nr 16/KS/6/2012/S/016).

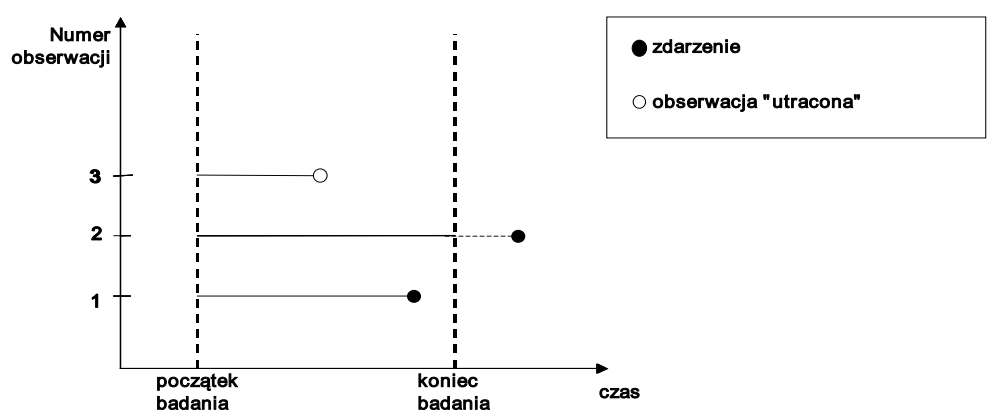
– utraty obserwowanej jednostki na skutek zdarzenia losowego (innego niż oczekiwane zdarzenie), którą przedwcześnie wyeliminowano z badania (np. zgon bezrobotnego szukającego pracy w analizie czasu pozostawania bez pracy, zniszczenie mieszkania przeznaczonego do sprzedaży (w analizie czasu sprzedaży mieszkań na rynku wtórnym) na skutek pożaru,
– zaprzestania procesu obserwacji badanej jednostki z powodu zakończenia badania, przy czym ewentualne zdarzenie związane z tą jednostką w późniejszym czasie nie zostanie już odnotowane (np. zaprzestanie obserwacji bezrobotnego, który do momentu zakończenia badania nie znalazł pracy; zakończenie obserwacji mieszkania, którego nie sprzedano do momentu zakończenia badania).

Zwróćmy uwagę, że w pierwszym przypadku cenzurowanie prawostronne ma charakter *losowy*, a w drugim ma charakter *ustalony*.

Przyjmijmy, że każda losowa jednostka ma swój własny czas trwania T, natomiast czas mierzony do momentu cenzurowania oznaczmy przez C (zakładamy też, że zmienne losowe T oraz C są niezależne dla każdej obserwowanej jednostki). Wówczas w przypadku losowego cenzurowania każdą jednostkę można scharakteryzować wektorem (Y, δ), gdzie:

$$Y = min(T, C),$$
$$\delta = I(T \leq C) = \begin{cases} 1, & gdy \quad T \leq C \\ 0, & gdy \quad T > C. \end{cases} \tag{1}$$

Zatem wskaźnik δ przyjmuje wartość 1, jeśli jednostka nie jest cenzurowana lub wartość 0, gdy jednostka jest cenzurowana.



Rysunek 1. Cenzurowanie prawostronne

Źródło: opracowanie własne na podstawie Balicki (2006).

Przykład cenzurowania prawostronnego przedstawiono na rysunku 1. Obserwację nr 3 należy traktować jako cenzurowaną prawostronnie, gdyż została utracona przed końcem badania, natomiast obserwacja nr 2 jest również cenzurowana prawostronnie, gdyż udało jej się „przetrwać" okres badania bez wystąpienia zdarzenia, które było przedmiotem badania. Jedynie w przypadku jednostki nr 1 zaobserwowano zdarzenie, które było przedmiotem badania w analizie historii zdarzeń i ta obserwacja cenzurowaną nie jest.

Przyjmijmy, że analizowane zdarzenie nastąpiło przed początkowym momentem procesu obserwacji, ale nie wiadomo dokładnie kiedy. Mówimy wówczas o cenzurowaniu lewostronnym. Z takim cenzurowaniem spotykamy się np. przy badaniu zapadalności na pewną chorobę w grupie pacjentów, spośród których niektórzy nie potrafią precyzyjnie określić kiedy w przeszłości wystąpiły u nich pierwsze objawy choroby. W sytuacji, gdy cenzurowanie jednostek następuje równocześnie przed momentem rozpoczęcia procesu obserwacji i po jego zakończeniu, to mówimy o cenzurowaniu obustronnym.

Jak już wspomniano przedmiotem badania w analizie trwania jest czas, jaki upływa od początku procesu obserwacji, aż do momentu zaistnienia zdarzenia kończącego dalszą obserwację jednostki (T), przy czym czas ten jest traktowany jako zmienna losowa o nieujemnych wartościach. Funkcja gęstości prawdopodobieństwa $f(t)$ tej zmiennej losowej pozwala określić rozkład liczby zdarzeń w czasie. Funkcję przeżycia (trwania) w analizie historii zdarzeń definiuje się następująco:

$$S(t) = 1 - F(t), \tag{2}$$

gdzie $F(t)$ oznacza dystrybuantę zmiennej losowej T ($F(t) = P(T < t)$). Funkcja przeżycia wyraża więc prawdopodobieństwo, że czas konieczny do zaistnienia zdarzenia przekroczy wartość t.

Istotnym zagadnieniem w analizie trwania jest testowanie, czy dwie populacje mają te same funkcje przeżycia. W literaturze można spotkać wiele testów do porównywania funkcji przeżycia. Część z nich doczekało się również oprogramowania w komputerowych pakietach statystycznych. Celem niniejszego opracowania jest ocena efektywności testów najczęściej stosowanych w analizie przeżycia. Cel ten osiągnięto poprzez badania symulacyjne. Godzi się zauważyć, że samo zagadnienie badania mocy testu krzywych przeżycia nie jest nowe i było już podejmowane w literaturze przedmiotu (por. Latta 1981; Magel 1991; Suciu i inni 2004; Jurkiewicz i Wycinka, 2011) w odniesieniu do wybranych testów statystycznych. W niniejszym artykule skupiono się natomiast na testach oprogramowanych w pakiecie komputerowym *STATISTICA*, w związku z tym wnioski z badań symulacyjnych mogą się przekładać na rekomendacje dla użytkowników tego pakietu statystycznego.

Badania symulacyjne przeprowadzono w środowisku R oraz programie *STATISTICA* za pomocą skryptu języka Visual Basic. W ramach symulacji generowano próby losowe z rozkładu Weibulla o wyróżnionych wartościach parametrów kształtu

i skali. Dla wybranej konfiguracji parametrów zdefiniowano rozkład referencyjny, z którym porównywane były pozostałe rozkłady. Wybór rozkładu Weibulla wynikał z tego, że jest to rozkład często stosowany w praktyce i jednocześnie dopuszczający szerokie spektrum postaci funkcji przeżycia czy funkcji hazardu. W badaniach symulacyjnych modyfikowano również rozmiary próby oraz odsetek obserwacji cenzurowanych. Za pomocą symulacji badano poziom błędu pierwszego rodzaju oraz moc porównywanych testów statystycznych służących testowaniu hipotezy zerowej głoszącej równość krzywych przeżycia w dwóch grupach.

2. TESTY SŁUŻĄCE DO PORÓWNANIA ROZKŁADÓW CZASU TRWANIA ZJAWISK

Do porównywania czasu trwania zjawisk służą testy istotności różnic pomiędzy dwoma krzywymi przeżycia. Weryfikujemy wówczas hipotezę zerową postaci $H_0 : S_1(t) = S_2(t)$. W zależności od problemu badawczego możemy rozpatrywać jedną z następujących hipotez alternatywnych: $H_1 : S_1(t) \neq S_2(t)$, $H_1 : S_1(t) > S_2(t)$ lub $H_1 : S_1(t) < S_2(t)$.

Testów do porównywania funkcji przeżycia w literaturze można spotkać wiele. Niestety, nie ma jednolitej konwencji w nazywaniu różnych statystyk testowych, a tym samym procedur na nich opartych, dlatego też Blossfeld, Golsch, Rohwer (2007, s. 79) omawiając procedury oprogramowane w pakiecie Stata, podają alternatywne nazwy pod jakimi można je spotkać w literaturze, jak i oprogramowaniu statystycznym. I tak na przykład procedurę log-rank zaimplementowaną w STATA można spotkać[2] w literaturze tematu pod nazwą Generalized Savage Test, a w pakiecie statystycznym BMDP pod nazwą Mantel-Cox log-rank. Z kolei procedura Wilcoxona-Breslowa-Gehana w programie STATA to[3] Generalized Wilcoxon (Breslow) w BMDP, a tylko Wilcoxon w SASie. Zdarza się niestety również tak, że pod tą samą nazwą w różnych pakietach statystycznych zaimplementowane są procedury istotnie różniące się algorytmicznie, a to może skutkować istotnie różnymi wynikami prowadzonych badań. Na przykład najpopularniejszą procedurę log-rank możemy spotkać w różnych wersjach, zarówno w dostępnym oprogramowaniu statystycznym, jak i w literaturze naukowej. Pyke i Thompson (1986) wyróżniają np. trzy testy log-rank: log-rank test wg Peto i Peto (por. Peto, Peto, 1972; Peto, Pike, 1973), test Coxa-Mantela (por. Cox, 1959, 1972; Mantel, 1966), test Mantela–Haenszela (por. Mantel, Heanszel, 1959).

W pakiecie *STATISTICA* dla użytkowników dostępnych jest pięć testów do porównywania dwóch funkcji przeżycia: test Wilcoxona wg Gehana, test Wilcoxona wg Peto i Peto, test log-rank, test F Coxa, test Coxa-Mantela. Biorąc pod uwagę fakt, iż w praktyce badawczej najczęściej wykorzystuje się dostępne, gotowe oprogramowanie, autorzy niniejszego opracowania zdecydowali się przeprowadzić badania symulacyjne, których celem jest porównanie efektywności testów służących do

[2] Ibidem.
[3] Ibidem.

porównywania czasu trwania zjawisk, w wersjach zaimplementowanych w pakiecie *STATISTICA*. Z powodu wspomnianego braku „jednolitej konwencji" w kwestii nazewnictwa testów badania dotyczące efektywności poprzedzone zostały szczegółową prezentacją algorytmów porównywanych testów w wersjach zaimplementowanych w pakiecie *STATISTICA*. Do opisu testów autorzy wykorzystywali zarówno oryginalne artykuły źródłowe w których opisane zostały wykorzystywane procedury, jak i wskazówki Stanisza (2007) dotyczące wersji zaimplementowanych w *STATISTICA* procedur.

Prezentację tych procedur rozpoczniemy od testów będących modyfikacjami nieparametrycznego testu Wilcoxona rangowanych znaków[4], służącego do porównywania rozkładów cech w dwóch populacjach, dla których obserwacje są zestawione w pary. Test Wilcoxona jest alternatywą dla testu par, gdy nie jest spełnione założenie o normalności rozkładów cech w obu populacjach. Gehan (1965) i bracia Peto (1972) zaproponowali modyfikacje testu Wilcoxona pod kątem stosowania go do badań czasu trwania zjawisk, gdy zazwyczaj w badaniu pojawiają się obserwacje cenzurowane (ucięte).

Test Wilcoxona wg Gehana

Na wstępie przyjmijmy za Gehanem (1965) następujące oznaczenia. Niech n_1 oznacza liczbę obserwacji w próbie pierwszej, a n_2 w próbie drugiej, przy czym:

- dla próby pierwszej: $x'_1, \ldots, x'_{s_1}$ to s_1 obserwacji cenzurowanych, a pozostałe $x_{s_1+1}, \ldots, x_{n_1}$ będą $(n_1 - s_1)$ – obserwacjami niecenzurowanymi;
- dla próby drugiej: $y'_1, \ldots, y'_{s_2}$ będą obserwacjami cenzurowanymi, natomiast $y_{s_2+1}, \ldots, y_{n_2}$ niecenzurowanymi w tej grupie.

Gehan (1965) opracował test wykorzystujący statystykę W zdefiniowaną następująco:

$$W = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij}, \tag{3}$$

gdzie wartość U_{ij} wyznaczana jest ze wzoru:

$$U_{ij} = \begin{cases} -1 & gdy\ x_i < y_j\ lub\ x_i \leq y'_j, \\ 0 & gdy\ x_i = y_j\ lub\ \left(x'_i, y'_j\right)^*\ lub\ x'_i < y_j\ lub\ y'_j < x_i, \\ 1 & gdy\ x_i > y_j\ lub\ x'_i \geq y_j. \end{cases} \tag{4}$$

---

[4] Niestety w polskiej literaturze statystycznej funkcjonuje niewłaściwa nazwa tego testu. *Wilcoxon signed-rank test* powinien być tłumaczony jako *Wilcoxona test znakowanych rang*. Popularna nazwa polska jest niepoprawna, bo znaki są efektem pomiaru w skali nominalnej i w żaden sposób nie można ich porangować (bo to pomiar w skali porządkowej).

* Obie obserwacje są cenzurowane.

Konstrukcja statystyki W nawiązuje (por. Gehan, 1965) do statystyk testowych: Wilcoxona, Manna-Whitney'a oraz do statystyki W Kendalla, w przypadku których nie występują obserwacje cenzurowane ani powiązane.

Gehan wykazał, że przy założeniu prawdziwości hipotezy zerowej, statystka W ma asymptotyczny rozkład normalny z wartością oczekiwaną $E(W) = 0$ i wariancją daną wzorem:

$$D^2(W) = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)}\left(\sum_{i=1}^{s} m_i M_{i-1}\,(M_{i-1}+1) + \sum_{i=1}^{s} l_i M_i\,(M_i+1) + \sum_{i=1}^{s} m_i\,(n_1+n_2-M_i-L_{i-1})(n_1+n_2-3M_{i-1}-m_i-L_{i-1}-1)\right), \quad (5)$$

gdzie $M_j = \sum_{i=1}^{j} m_i,\ M_0 = 0,\ L_j = \sum_{i=1}^{j} l_i,\ L_0 = 0,$

m_i – liczba obserwacji niecenzurowanych w i-tej randze w uporządkowaniu rangowym obserwacji nieuciętych z różnymi wartościami,

l_i – liczba obserwacji prawostronnie cenzurowanych o wartościach większych niż obserwacja o randze i ale mniejszych niż obserwacja o randze $(i + 1)$,

s – łączna liczba obserwacji niecenzurowanych o różnych czasach przeżycia.

Mantel (1967) zaproponował znaczne uproszczenie skomplikowanej obliczeniowo procedury Gehana. W oryginalnej procedurze Gehana każda obserwacja z próby pierwszej jest porównywana ze wszystkimi obserwacjami z drugiej próby. Mantel (1967) zaproponował, by obydwie próby połączyć w jedną o $(n_1 + n_2)$ obserwacjach i każdą obserwację porównywać z pozostałymi $(n_1 + n_2 – 1)$ obserwacjami. Wartości U_i $(i = 1, \ldots, n_1 + n_2)$ wyznaczane są jako różnice pomiędzy liczbą pozostałych spośród $(n_1 + n_2 – 1)$ obserwacji definitywnie mniejszych od i-tej obserwacji, a liczbą tych spośród nich, które są definitywnie większe od tej obserwacji (por. np. Mantel, 1967; Lininger i in., 1979).

Mantel wykazał, że statystykę W Gehana można zapisać jako sumę wartości U_i odpowiadających pierwszej próbie. Wykazał również, że statystyka W ma asymptotyczny rozkład normalny o wartości oczekiwanej $E(W) = 0$ i wariancji:

$$D^2(W) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} U_i^2}{(n_1+n_2)(n_1+n_2-1)}. \quad (6)$$

Mantel (1967) podał również prosty algorytm wyznaczania wartości U_i.

Algorytm ten polega na uporządkowaniu obu połączonych prób w porządku niemalejącym.

Wyznaczamy wartości R_{1i} przeprowadzając następujące kroki:

1. Przyporządkowujemy rangi obserwacjom od najmniejszej do największej, pomijając obserwacje prawostronnie cenzurowane.
2. Każdej obserwacji prawostronnie cenzurowanej przyporządkowujemy następną wyższą rangę.
3. Niecenzurowanym obserwacjom powiązanym (o tej samej wartości) przyporządkowujemy najmniejszą wartość rangi z przypisanych do tych obserwacji w kroku 1.
4. Redukujemy wartości rang dla obserwacji lewostronnie uciętych do 1.

Następnie wyznaczamy wartości R_{2i}:

1. Przyporządkowujemy rangi obserwacjom od wartości największej do najmniejszej, pomijając obserwacje lewostronnie cenzurowane.
2. Każdej obserwacji lewostronnie cenzurowanej przyporządkowujemy następną wyższą rangę.
3. Niecenzurowanym obserwacjom powiązanym (o tej samej wartości) przyporządkowujemy najmniejszą wartość rangi z przypisanych do tych obserwacji w kroku 1.
4. Redukujemy wartości rang dla obserwacji prawostronnie uciętych do 1.

Następnie wyznaczamy wartości statystyki $U_i = R_{1i} - R_{2i}$ dla $i = 1, \ldots, n_1 + n_2$.

Algorytm ten został zaimplementowany w pakiecie *STATISTICA* w wersji uproszczonej dla cenzurowania prawostronnego (por. Stanisz, 2007).

W pakiecie *STATISTICA* podana jest wartość empiryczna statystyki testowej[5]: $\frac{W-0,5}{D(W)}$ mającej asymptotyczny standaryzowany rozkład normalny (Stanisz, 2007).

Test Wilcoxona wg Peto i Peto

W 1972 roku bracia Peto (1972) zaproponowali modyfikację testu Wilcoxona opartą na ocenach funkcji przeżycia wyznaczonych metodą Kaplana-Meiera dla połączonych prób.

Algorytm (por. Peto, Peto, 1972; Lee, Desu, Gehan, 1975) polega na przyporządkowaniu obserwacjom niecenzurowanym t_i wartości $S(t_i) + S(t_{i-1}) - 1$, natomiast wartościom cenzurowanym prawostronnie t_i wartości $S(t_k) - 1$ wyznaczonej dla największej nieuciętej wartości $t_k \leq t_i$, gdzie S jest estymatorem funkcji przeżycia

[5] We wzorze statystyki testowej uwzględniono dyskretny charakter statystyki W (wprowadzono korektę na ciągłość).

Kaplana-Meiera dla obu połączonych prób. Suma przyporządkowanych w ten sposób wartościdla obu prób wynosi 0.

Algorytm procedury zaimplementowanej w programie *STATISTICA* jest następujący:

1. Wyznaczamy oceny funkcji przeżycia metodą Kaplana-Meiera dla połączonych prób:
   $S(t_i)$ – ocena funkcji przeżycia dla *i*-tej obserwacji niecenzurowanej.
2. Porządkujemy niemalejąco czasy przeżycia obserwacji niecenzurowanych oraz czasy obserwacji jednostek cenzurowanych dla obu prób łącznie. Najmniejszej obserwacji niecenzurowanej przyporządkowujemy odpowiadającą jej ocenę funkcji przeżycia, a następnie poszczególnym czasom przeżycia t_i przyporządkowujemy wartości W_i wyznaczone ze wzoru:

$$W_i = \begin{cases} S(t_i) + S(t_{i-1}) - 1 & \text{dla } obserwacji\ niecenzurowanej\ t_i, \\ S(t_k) - 1 & \text{dla } obserwacji\ cenzurowanej\ t_i,\ dla\ której\ t_k\ jest \\ & największą\ obserwacją\ niecenzurowaną\ taką, że\ t_k \leq t_i. \end{cases} \quad (7)$$

W programie *STATISTICA* powtarzającym się wartościom obserwacji niecenzurowanych ostatecznie przyporządkowywana jest średnia arytmetyczna odpowiadających im wartości W_i.

Przy założeniu prawdziwości hipotezy zerowej, statystyka *WW* będąca sumą wartości W_i dla próby pierwszej, ma rozkład asymptotycznie normalny z wartością oczekiwaną $E(WW) = 0$ i wariancją daną wzorem:

$$D^2(WW) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} W_i^2}{(n_1 + n_2)(n_1 + n_2 - 1)}. \quad (8)$$

W pakiecie *STATISTICA* podawana jest wartość empiryczna statystyki testowej: $\frac{WW}{D(WW)}$ mającej asymptotyczny standaryzowany rozkład normalny.

Test log-rank

Test log-rank został zaproponowany w 1966 roku przez Mantela (1966), ale nazwę tego testu zaproponowali bracia Peto (1972). Konstrukcja tego testu opiera się na logarytmie funkcji przeżycia. W literaturze tematu można spotkać różne wersje i modyfikacje oryginalnego testu log-rank Mantela np. test Mantela-Haenszela log-rank, test Coxa-Mantela logrank czy też test log-rank Peto i Peto. Poniżej przedstawimy wersję testu zaproponowaną przez Mantela (1966) z wykorzystaniem estymatora logarytmu funkcji przeżycia zaproponowanego przez Altshulera (1970). Wersja ta została zaim-

plementowana w programie *STATISTICA* pod nazwą test log-rank, a jej algorytm jest następujący (Lee, Desu Gehan, 1975, Stanisz 2007):

1. Porządkujemy rosnąco czasy przeżycia obserwacji niecenzurowanych oraz czasy obserwacji jednostek cenzurowanych dla obu prób łącznie:

$$t_{(1)} < t_{(2)} < \ldots < t_{(l)},$$

   a następnie przyporządkowujemy czasom przeżycia $t_{(i)}$ odpowiadającą im liczbę zdarzeń końcowych (niepożądanych) $m_{(i)}$.
2. W kolejnym kroku czasom przeżycia $t_{(i)}$ przyporządkowujemy $r_{(i)}$ liczbę obserwacji w zbiorze ryzyka[6] $R(t_{(i)})$, czyli liczbę obserwacji kompletnych oraz cenzurowanych z obu prób o czasie przeżycia nie krótszym niż $t_{(i)}$.
3. Logarytm funkcji przeżycia estymujemy metodą, do której rozpowszechnienia przyczynili się bracia Peto (1972), a zaproponowaną[7] przez Altshulera (1970):

$$e\left(t_{(i)}\right) = \sum_{j \le t_{(i)}} \frac{m_{(j)}}{r_{(j)}}. \tag{9}$$

4. W kolejnym kroku wyznaczane są wartości W_i zgodnie ze wzorem:

$$W_i = \begin{cases} 1 - e(t_{(i)}) & \text{dla obserwacji niecenzurowanej } t_{(i)}, \\ -e(t_{(k)}) & \text{dla obserwacji cenzurowanej } t_{(i)}, \text{ dla której } t_{(k)} \text{ jest} \\ & \text{największą obserwacją niecenzurowaną taką, że } t_{(k)} \le t_{(i)}. \end{cases} \tag{10}$$

   Zauważmy, że większym wartościom obserwacji niecenzurowanych odpowiadają mniejsze wartości W_i, natomiast obserwacje cenzurowane mają przyporządkowane ujemne wartości W_i. Suma W_i dla wszystkich obserwacji wynosi 0.
5. Test log-rank jest oparty na sumie wartości W_i dla jednej z dwóch prób (por. Lee, Desu, Gehan, 1975). W pakiecie *STATISTICA* w podsumowaniu wyników testu log-rank podawana jest suma *WW* wartości W_i dla obserwacji próby pierwszej. Przy założeniu prawdziwości hipotezy zerowej, statystyka *WW* ma rozkład taki jak w teście Wilcoxona wg Peto i Peto, czyli asymptotycznie normalny z wartością oczekiwaną zero i wariancją daną wzorem (8).
6. W pakiecie *STATISTICA* podawana jest wartość empiryczna statystyki: $\frac{WW}{D(WW)}$ mającej asymptotyczny standaryzowany rozkład normalny.

---

[6] Zbiór wszystkich kompletnych oraz cenzurowanych obserwacji o czasach przeżycia nie krótszych od t nazywany jest w literaturze zbiorem ryzyka w chwili t i oznaczany $R(t)$ (patrz np. Lee, Desu, Gehan, 1975).

[7] Metoda zaproponowana przez Altshulera (1970) dla danych prawostronnie cenzurowanych.

Test Coxa-Mantela

W 1972 roku Cox (1972) zaprezentował procedurę porównywania dwóch krzywych przeżycia. W literaturze tematu procedura ta spotykana jest zarówno pod nazwą testu Coxa (por. Desu, Lee, Gehan, 1975), jak i testu Coxa-Mantela. W pakiecie *STATISTICA* procedura Coxa (1972) występuje pod nazwą testu Coxa-Mantela.

Algorytm procedury Coxa-Mantela jest następujący (Cox, 1972, Desu, Lee, Gehan, 1975, Stanisz, 2007):

1. Porządkujemy rosnąco czasy przeżycia (obserwacji kompletnych) dla obu prób łącznie:

$$t_{(1)} < t_{(2)} < \ldots < t_{(k)},$$

a następnie przyporządkowujemy wyróżnionym czasom przeżycia $t_{(i)}$ odpowiadającą im liczbę zdarzeń końcowych (niepożądanych) $m_{(i)}$. Zauważmy, że:

$$\sum_{i=1}^{k} m_{(i)} = n_1+n_2 - (s_1 + s_2), \tag{11}$$

s_i – jest to liczba prawostronnie cenzurowanych obserwacji w i-tej próbie ($i = 1, 2$).

2. Następnie czasom przeżycia $t_{(i)}$ przyporządkowujemy $r_{(i)}$ liczbę obserwacji w zbiorze ryzyka $R(t_{(i)})$, czyli liczbę obserwacji kompletnych oraz cenzurowanych z obu prób o czasie przeżycia nie krótszym niż $t_{(i)}$.
3. Następnie zaobserwowanym czasom przeżycia $t_{(i)}$ przyporządkowujemy $A_{(i)}$ stosunek $r_{2,(i)}$ – liczby obserwacji z drugiej próby o czasie przeżycia nie krótszym niż $t_{(i)}$ oraz $r_{(i)}$:

$$A_{(i)} = \frac{r_{2,(i)}}{r_{(i)}}. \tag{12}$$

4. Wyznaczamy wartości statystyk:

$$U = n_2 - s_2 - \sum_{i=1}^{k} m_{(i)} A_{(i)}, \tag{13}$$

$$I = \sum_{i=1}^{k} \frac{m_{(i)}\left(r_{(i)} - m_{(i)}\right)}{r_{(i)} - 1} A_{(i)}\left(1 - A_{(i)}\right). \tag{14}$$

5. Następnie wyznaczamy wartość statystyki testowej C:

$$C = \frac{U}{\sqrt{I}}\,. \tag{15}$$

Cox (1972) wykazał, że przy założeniu prawdziwości hipotezy zerowej statystyka testowa C ma asymptotyczny rozkład normalny standaryzowany $N(0,1)$.

Test F Coxa

Test F Coxa został opisany przez Coxa (1964). Konstrukcja testu oparta jest na rozkładzie wykładniczym z parametrem $\lambda = 1$. Załóżmy, że obie próby pochodzą z rozkładu wykładniczego ($\lambda = 1$) o liczebnościach odpowiednio n_1, n_2, w tym odpowiednio s_1, s_2 jest liczbą obserwacji cenzurowanych. Testowanie przebiega w następujących etapach:

1. Porządkujemy niemalejąco czasy przeżycia obserwacji niecenzurowanych dla obu prób łącznie.
2. Kolejnym obserwacjom niecenzurowanym przyporządkowujemy wartości:

$$t_{rn} = \frac{1}{n} + \cdots + \frac{1}{n-r+1} \qquad \text{dla } r = 1, \ldots, (n_1 - s_1) + (n_2 - s_2), \tag{16}$$

gdzie $n = n_1 + n_2$.

W przypadku powtarzających się obserwacji niecenzurowanych, odpowiadające im wartości t_{rn} zastępujemy średnią arytmetyczną tych wartości.

Wartościom cenzurowanym przyporządkowujemy wartość:

$$t_{((n_1-s_1)+(n_2-s_2)+1),\, n} = \frac{1}{n} + \cdots + \frac{1}{s_1 + s_2}\,. \tag{17}$$

3. Następnie dla obu prób wyznaczamy średnie wartości $\bar{t}_1, \bar{t}_2$ sumując przyporządkowane w kroku 2 wartości t dla odpowiedniej próby, a następnie dzieląc przez liczbę niecenzurowanych obserwacji w tej próbie, czyli odpowiednio przez $(n_1 - s_1), (n_2 - s_2)$.

Cox (1964) pokazał, że iloraz tak wyznaczonych średnich $\bar{t}_1, \bar{t}_2$, przy założeniu prawdziwości hipotezy zerowej, ma rozkład $\boldsymbol{F}$ o $(2(n_1 - s_1), 2(n_2 - s_2))$ stopniach swobody.

3. WYNIKI EKSPERYMENTÓW SYMULACYJNYCH BADANIA MOCY TESTÓW PORÓWNYWANIA PRZEŻYĆ

Za pomocą symulacji badano poziom błędu pierwszego rodzaju oraz moc testów statystycznych opisanych w rozdziale 2. Symulacje przeprowadzono w środowisku statystycznym R oraz programie *STATISTICA* 10 za pomocą skryptu języka Visual Basic. W ramach symulacji generowano próby losowe z rozkładu Weibulla o parametrze kształtu równym 2 natomiast parametr skali przyjmował wartości 25, 30, 35, 40, 45 oraz 50. Wartość 50 definiowała rozkład referencyjny, czyli rozkład z którym porównywane były pozostałe rozkłady, włącznie z rozkładem referencyjnym. Przyjęte wartości parametru skali skutkowały następującymi wartościami median czasu przeżycia: 20,8; 25; 29,1; 33,3; 37,5; 41,6 miesięcy. Obserwacje cenzurowane generowano zgodnie z metodologią przedstawioną w Halabi i Singh (2004). Kolejnymi dwoma modyfikowanymi parametrami symulacji były rozmiar próby (równy dla obu porównywanych grup) oraz odsetek obserwacji cenzurowanych. Zestawienie scenariuszy symulacyjnych przedstawia tabela 1, natomiast rysunek 2 demonstruje porównywane krzywe przeżycia (krzywa najwyżej położona jest krzywą referencyjną).

Tabela 1.

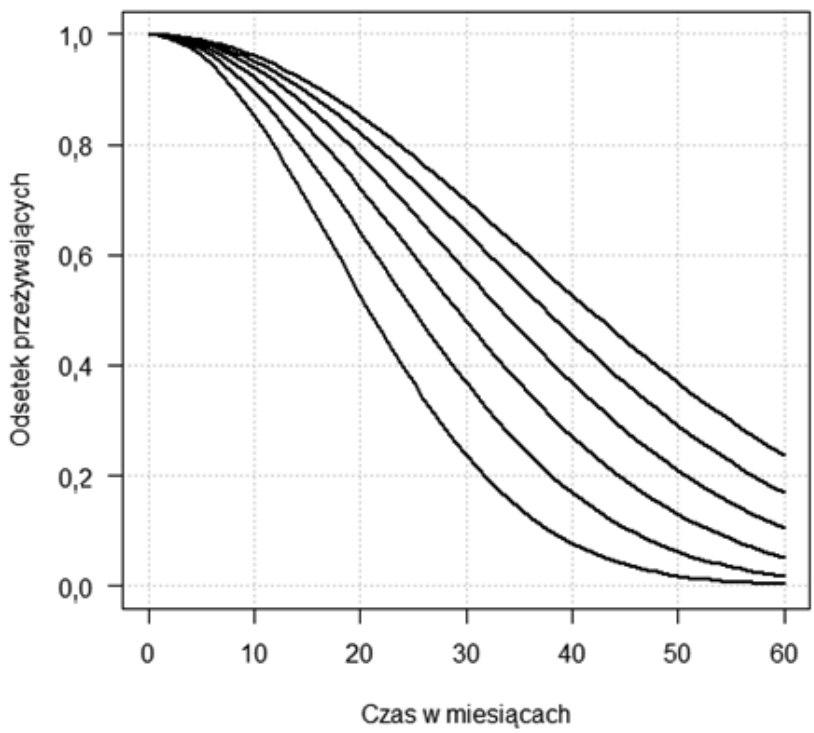
Schematy eksperymentów symulacyjnych

Rozmiary prób	Odsetek obserwacji cenzurowanych w próbie referencyjnej	Odsetek obserwacji cenzurowanych w próbie porównywanej	Parametr skali rozkładu Weibulla	Mediana w rozkładzie Weibulla o zadanej wartości parametru skali
50 100 300	0,3 0,5 0,7 0,3 0,3	0,3 0,5 0,7 0,5 0,7	25	20,8
			30	25,0
			35	29,1
			40	33,3
			45	37,5
			50	41,6

Źródło: ustalenia własne.

W rezultacie wstępnej analizy wyników symulacji okazało się, że wyniki dla testu F Coxa zdecydowanie odbiegają od wyników pozostałych testów. Nawet dla stosunkowo odległych krzywych przeżyć (mediany różniące się o około 21 miesięcy, czyli około dwukrotnie) moc tego testu wyniosła około 0,40, podczas gdy pozostałe testy osiągnęły moc równą 1 (czyli zawsze odrzucały nieprawdziwą hipotezę zerową). Podobne wyniki testu F Coxa, odbiegające od pozostałych, uzyskano również dla innych rozważanych liczebności prób, oraz procentów obserwacji cenzurowanych.

W literaturze tematu test F Coxa jest zalecany (por. np. Lee, Desu, Gehan, 1975[8]; Stanisz, 2007), w sytuacji, gdy próby są małe (mniejsze od 50), a obserwacji cenzurowanych nie ma lub jest ich bardzo niewiele. W przeprowadzonych badaniach symulacyjnych taka sytuacja nie była rozważana. Słabe wyniki testu F Coxa wynikają zapewne również z faktu, iż konstrukcja tego testu oparta jest na rozkładzie wykładniczym (z parametrem $\lambda = 1$), a w badaniu symulacyjnym próby generowane były z rozkładu Weibulla z parametrem kształtu wynoszącym 2 (rozkład Weibulla przechodzi w rozkład wykładniczy dla parametru kształtu równego jeden). Zatem test F Coxa został usunięty z dalszych rozważań.



Rysunek 2. Teoretyczne krzywe przeżyć wykorzystywane w eksperymentach symulacyjnych
Źródło: opracowanie własne.

Po kilku eksperymentach z różnymi funkcjami stwierdzono, że satysfakcjonującą aproksymację funkcji mocy testu daje funkcja logistyczna określona następującym wzorem:

[8] Test F Coxa (1964) występuje w artykule u Lee, Desu, Gehana (1975) pod nazwą testu F.

$$P(d) = \frac{1}{1 + b \cdot \exp(-c \cdot d)}, \tag{18}$$

gdzie d jest odległością pomiędzy medianami porównywanych krzywych przeżyć (miara odstępstwa od prawdziwości hipotezy zerowej). Ponadto zakłada się, że moc testu w warunkach prawdziwości hipotezy zerowej wynosi 0,05, stąd wynika wartość b równa 19. Naturalny w rozważanym zagadnieniu poziom nasycenia funkcji logistycznej wynosi 1 (licznik wyrażenia 18). Teoretyczna funkcja mocy testu była szacowana nieliniowym algorytmem Levenberga-Marquardta z wykorzystaniem kryterium najmniejszych kwadratów.

W związku z przyjętymi założeniami dotyczącymi funkcji aproksymującej (α=0,05 przy prawdziwości hipotezy zerowej oraz poziom nasycenia równy 1), w funkcji (18) pozostaje tylko jeden parametr (c). Im moduł tego parametru jest większy tym wykres funkcji mocy testu położony jest „wyżej". W związku z tym podzielono wartość c dla poszczególnych testów przez wartość c dla testu najlepszego (z największą wartością c). Otrzymujemy w ten sposób pewną miarę względnej mocy testu. Zestawienia relatywnej mocy porównywanych testów przy różnych liczbach obserwacji przedstawiono w tabelach 2–4.

Tabela 2.

Relatywna moc testów porównywania przeżyć dla n=50

Test	Proporcje obserwacji cenzurowanych (w procentach)				
	30/30	50/50	70/70	30/50	30/70
Gehana-Wilcoxona	0,939	0,875	0,903	0,875	0,841
Coxa-Mantela	1,000	1,000	1,000	1,000	1,000
Peto-Peto- Wilcoxona	0,968	0,925	0,903	0,907	0,808
Log-rank	0,977	0,953	0,916	0,891	0,618

Źródło: obliczenia własne.

Tabela 3.

Relatywna moc testów porównywania przeżyć dla n=100

Test	Proporcje obserwacji cenzurowanych (w procentach)				
	30/30	50/50	70/70	30/50	30/70
Gehana-Wilcoxona	0,883	0,889	0,897	0,902	0,880
Coxa-Mantela	1,000	1,000	1,000	1,000	1,000
Peto-Peto- Wilcoxona	0,929	0,936	0,949	0,941	0,880
Log-rank	0,983	0,985	0,932	0,918	0,732

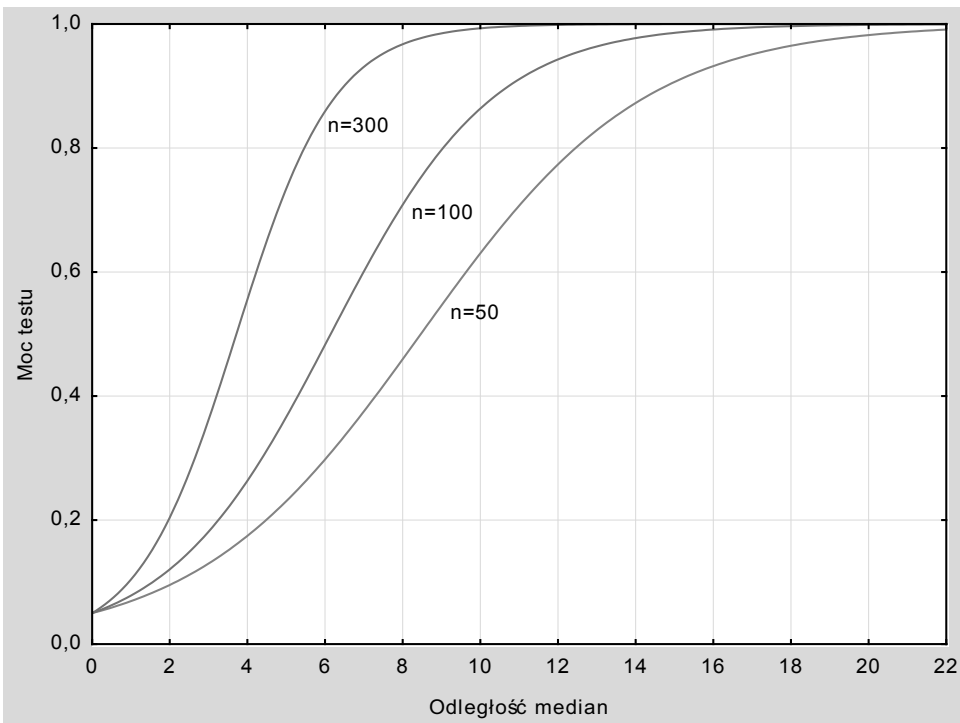
Źródło: obliczenia własne.

Tabela 4.

Relatywna moc testów porównywania przeżyć dla n=300

Test	Proporcje obserwacji cenzurowanych (w procentach)				
	30/30	50/50	70/70	30/50	30/70
Gehana-Wilcoxona	0,889	0,906	0,890	0,868	0,904
Coxa-Mantela	1,000	1,000	1,000	1,000	1,000
Peto-Peto- Wilcoxona	0,927	0,966	0,965	0,912	0,910
Log-rank	0,990	0,994	0,985	0,932	0,808

Źródło: obliczenia własne.



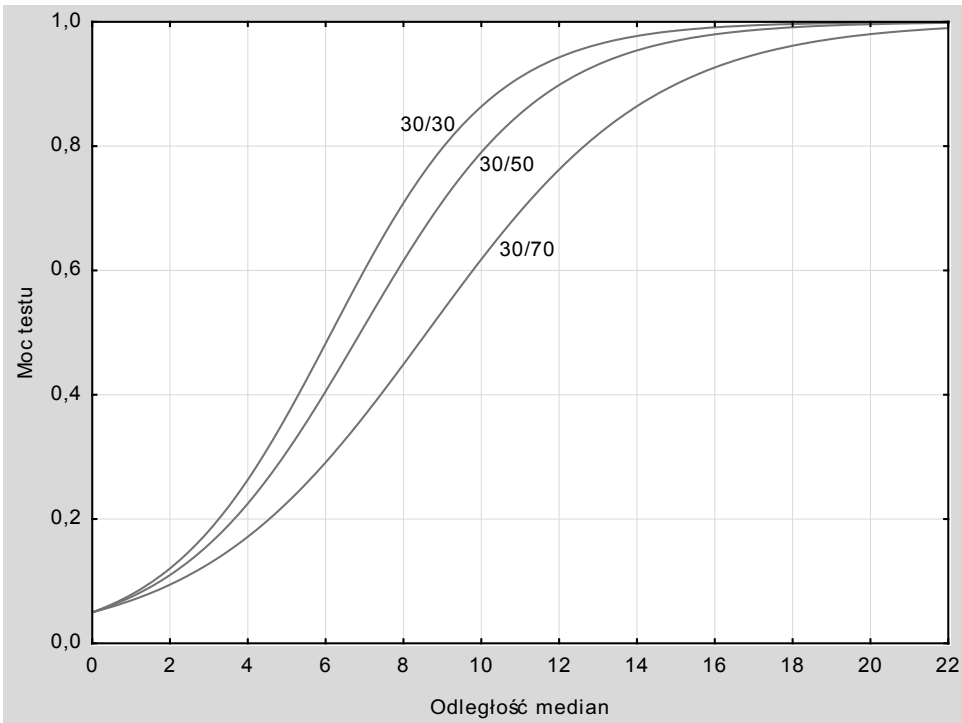
Rysunek 3. Porównanie mocy testu Cox-Mantela w zależności od liczebności próby (cenzurowanie w obu próbach 30%)

Źródło: opracowanie własne.

Wyniki symulacji przynoszą jednoznaczne rozstrzygnięcie. Dla wszystkich schematów eksperymentów symulacyjnych najlepszy okazał się test Coxa-Mantela. Na drugim miejscu plasował się na ogół test long-rank. Różnica mocy między tymi

testami zależała od proporcji obserwacji cenzurowanych. Przy małym ich procencie obydwa testy mają praktycznie porównywalną moc. Test log-rank okazał się bardzo wrażliwy na nierównomierną proporcję cenzurowania, szczególnie przy małej liczbie obserwacji kompletnych w stosunku do rozmiaru próby. W konsekwencji dodatkowe analizy zostaną zaprezentowane tylko dla testu Coxa-Mantela (rysunki 3–5).

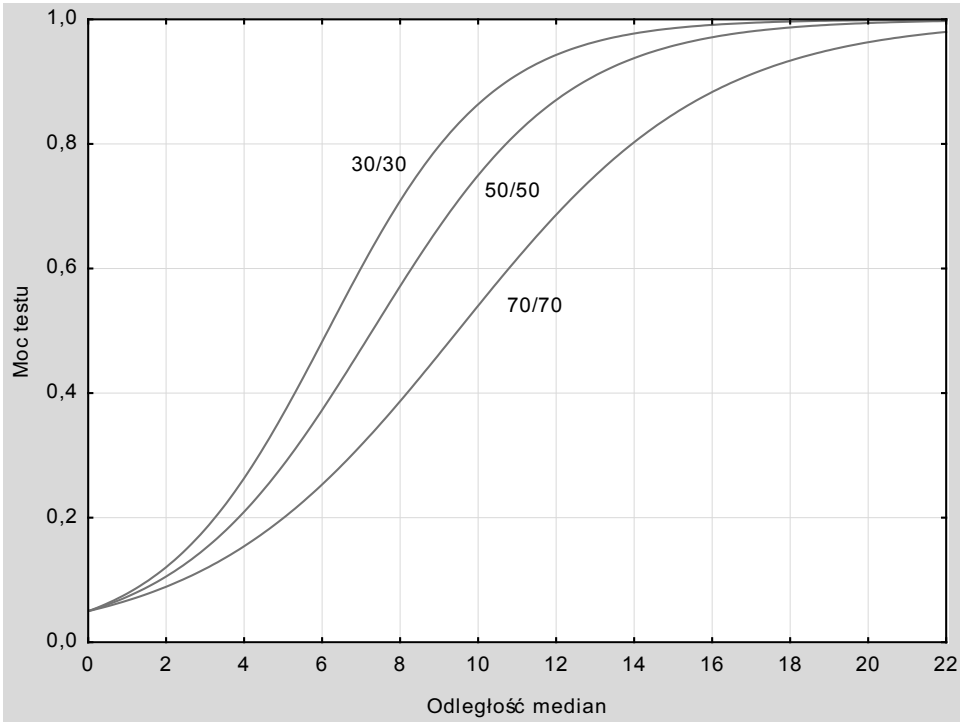
Na rysunku 3 widać oczywistą reakcję testu na zwiększanie liczebności próby – moc testu rośnie w miarę wzrostu liczebności próby. Do dalszej, bardziej szczegółowej prezentacji wyników wybrano liczebność próby równą 100.



Rysunek 4. Moc testu Coxa-Mantela przy rosnącej proporcji obserwacji cenzurowanych w drugiej próbie (C1/C2), n = 100

Źródło: opracowanie własne.

Moc testu Coxa-Mantela jest najwyższa przy wyrównanej proporcji obserwacji cenzurowanych w obu próbach. Jeżeli udział cenzurowanych staje się coraz bardziej różny, to moc testu słabnie. Zwiększanie się udziału obserwacji cenzurowanych pogarsza moc testu również w przypadku równej proporcji takich obserwacji w obydwu próbach (rysunek 5).



Rysunek 5. Moc testu Coxa-Mantela w zależności od proporcji obserwacji cenzurowanych (C1/C2), n=100

Źródło: opracowanie własne.

4. Podsumowanie

Zwiększanie liczebności próby polepsza moc zbadanych testów, co jest elementarnym wymaganiem stawianym każdemu testowi statystycznemu. Zwiększanie udziału obserwacji cenzurowanych w generowanych próbach pogarsza moc testów. W podsumowaniu można stwierdzić, że test Coxa-Mantela okazał się w przeprowadzonych badaniach nieco lepszy od zdecydowanie najpopularniejszego w praktycznych zastosowaniach testu log-rank. Różnice mocy między tymi testami stają się znaczące dopiero przy dużym procencie obserwacji cenzurowanych, szczególnie gdy cenzurowanie nie jest równomierne w porównywanych próbach. Inny ważny wniosek z badań to stwierdzenie, że test F Coxa zaimplementowany w programie *STATISTICA* powinien być stosowany z ostrożnością, gdyż nie jest to test uniwersalny.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

[1] Altshuler B., (1970), Theory for the Measurement of Competing Risks in Animal Experiments, *Math. Biosciences*, 6, 1-11.
[2] Balicki A., (2006), *Analiza przeżycia i tablice wymieralności*, PWE, Warszawa.
[3] Blossfeld H-P., Golsch K., Rohwer G., (2007), *Event History Analysis with Stata*, New Jersey: Lawrence Erlbaum Associates.
[4] Cox D. R., (1959), The Analysis of Exponentially Distributed Life-times with Two Types of Failure, *Journal of the Royal Statistical Society*, Series B (Methodological), 21, 411-421.
[5] Cox D. R., (1964), Some Applications of Exponential Ordered Scores, *Journal of the Royal Statistical Society*, Series B (Methodological), 26 (1), 103-110.
[6] Cox D. R., (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society*, Series B (Methodological), 34 (2), 187-220.
[7] Deszczyńska A., (2011), *Model hazardów proporcjonalnych Coxa*, *Matematyka stosowana*, tom, 13/54, Instytut Matematyczny PAN, Warszawa.
[8] Gehan, E. A., (1965), A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-censored Samples, *Biometrika*, 52, 203-223.
[9] Halabi S., Singh B., (2004), Sample Size Determination for Comparing Several Survival Curves with Unequal Allocations, *Statistics in Medicine*, 23, 1793-1815.
[10] Jurkiewicz T., Wycinka E., (2011), Significance Tests of Differences Between Two Crossing Survival Curves for Small Samples, w: Domański Cz., Zielińska-Sitkiewicz K. (red.), *Methodological Aspects of Multivariate Statistical Analysis, Statistical Models and Applications*, Łódź.
[11] Latta R. B., (1981), Monte Carlo Study of Some Two-Sample Rank Tests With Censored Data, *Journal of Statistical Association*, 76 (375), 713-719.
[12] Lee E.T., Desu M. M., Gehan E. A., (1975), A Monte Carlo Study of the Power of Some Two-sample Tests, *Biometrika*, 62 (2), 425-432.
[13] Lininger L., Gail M. H., Green S. B., Byar D. P., (1979), Comparison of Four Tests for Equality of Survival Curves in the Presence of Stratification and Censoring, *Biometrika*, 66 (3), 419-428.
[14] Magel R. C., (1991), Estimating the Power of the Gehan Test, *Biometrical Journal*, 88 (8), 985-997.
[15] Mantel N., (1966), Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration, *Cancer Chemotherapy Reports*, 50, 163-170.
[16] Mantel N., (1967), Ranking Procedure for Arbitrarily Restricted Observation, *Biometrics*, 23 (1), 65-78.
[17] Mantel N., Haenszel W., (1959), Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease, *Journal of the National Cancer Institute*, 22, 719-748.
[18] Peto R., Peto J., (1972), Asymptotically Efficient Rank Invariant Procedures, *Journal of the Royal Statistical Society*, Series A (General), 135 (2), 185-207.
[19] Peto R., Pike M. C., (1973), Conservatism of the Aproximation in the Logrank Test for Survival Data or Tumor Incidence Data, *Biometrics*, 29 (3), 579-84.
[20] Pyke D. A., Thompson J. N., (1986), Statistical Analysis of Survival and Remotal Rate Experiments, *Ecology*, 67 (1), Ecological Society of America, 240-245.
[21] http://bio.research.ucsc.edu/people/thompson/PublPDFs/025Statistical.pdf
[22] Stanisz A., (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, t. 3, StatSoft.
[23] Suciu G. P., Lemeshow S., Moeschberger M., (2004), *Statistical Tests of the Equality of Survival Curves: Reconsidering the options*, w: Balakrishnan N., Rao C. R. (red.), *Handbook of Statistics 23*, *Advances in Survival Analysis,* Elsevier B. V.

ANALIZA MOCY WYBRANYCH TESTÓW JEDNORODNOŚCI CZASÓW TRWANIA DLA POPULACJI O ROZKŁADZIE WEIBULLA

Streszczenie

W ostatnim latach testy porównywania czasu trwania zjawisk znajdują coraz więcej zastosowań w analizie zagadnień ekonomicznych. Przykładami może być analiza czasu pozostawania na bezrobociu, czasu poszukiwania pracy, czasu istnienia przedsiębiorstwa, itp.

W literaturze można spotkać wiele testów do porównywania funkcji przeżycia. Autorzy niniejszego opracowania zdecydowali się przeprowadzić badania symulacyjne, których celem jest porównanie efektywności najczęściej stosowanych testów służących do porównywania czasu trwania zjawisk, w wersji zaimplementowanej w pakiecie *STATISTICA*. Za pomocą symulacji badano poziom błędu pierwszego rodzaju oraz moc następujących testów statystycznych służących testowaniu hipotezy zerowej głoszącej równość krzywych przeżycia w dwóch populacjach. Analizie poddano następujące testy: Wilcoxona wg Gehana, F Coxa, Coxa-Mantela, Wilcoxona wg Peto i Peto i log-rank. W ramach symulacji generowano próby losowe z rozkładu Weibulla.

Słowa kluczowe: funkcja przeżycia, porównywanie funkcji przeżycia, badania symulacyjne

THE POWER ANALYSIS OF TESTS FOR COMPARING SURVIVAL OF WEIBULL DISTRIBUTED POPULATIONS

Abstract

Recently, tests for comparing survival distributions become more and more popular and used in applied economics. Unemployment duration, time needed to find a new job, enterprise survival or waiting for a commodity to be sold are good examples. There are a number of tests to compare survival distributions proposed in statistical literature. The aim of this research was to analyze, by the means of computer simulations, the effectiveness of survival tests as implemented in *STATISTICA* software. The following tests have been outlined and compared: Wilcoxon, Geehan, Cox-Mantel, Peto & Peto and log rank. Random samples were generated from Weibull distribution.

Keywords: survival function, comparing survival, Monte Carlo research

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

PAWEŁ STRAWIŃSKI

SMALL SAMPLE PROPERTIES OF MATCHING WITH CALIPER[1]

1. INTRODUCTION

Matched sampling methodology is used to mimic control experiments and estimate causal treatment effects. Unlike in the experimental sciences, in the social sciences fully controlled experiments usually are not possible to conduct due to various reasons. This methodology can be applied in all situations when one has a treatment, for example particular social policy, a group of treated individuals and a group of untreated individuals. The ultimate goal is to measure the average difference in outcome between treated and non-treated individuals, so called, the average treatment effect. The problem is that participants and non-participants usually differ in observed pre-treatment characteristics X. The second problem arises because an individual can be treated or non-treated but not both at the same time.

The role of matching is to adjust a non-treated group in order to make it comparable with a treated group. It allows us to alter the differences in the distributions of characteristics between a treated and a non-treated group. Even when the differences are observed it is necessary to adjust for those differences to obtain unbiased estimates of treatment effects. Matching estimators link units from a treated group with those from a non-treated group. Matching is typically done without replacement, so each non-treated observation is used as a match only once and matches are independent (Abadie, Imbens, 2011). The procedure seeks for each treated observation a non-treated observation, also called control observation, identical or very similar to it in observed characteristics. The most widely used types of matching are Mahalanobis distance matching, propensity score matching and kernel based matching. Among them the propensity score matching plays a fundamental role, since it reduces the curse of dimensionality problem and allows for one dimensional non-parametric regression (Rosenbaum, Rubin, 1983). The propensity score itself is a probability of receiving a treatment. It is also important to mention that matching on the propensity score is sufficient to balance the observed covariates. More recently developed nonparametric matching estimators described, for example, in Heckman et al. (1997, 1998) use

[1] This research is partly financed by the Ministry of Science and Higher Education grant N111 109335 (50%) and the Faculty of Economic Sciences Warsaw University grant (50%).

weighted average over multiple observations to construct matches. However, those methods are computationally demanding.

In this paper we focus on a particular matching technique that is, matching with caliper and its applications in small samples. The caliper mechanism has been proposed to control for matching quality and prevent from inexact matches. There exists a practical trade-off when obtaining matched samples between desires to (i) find matches for all treated units and (ii) use only matched treated-control pairs that are extremely similar to each other (Rosenbaum, Rubin, 1985b). The former situation leads to inexact matching, as matched objects may differ substantially, while the latter leads to incomplete matching.

The main motivation is that the finite sample properties of caliper mechanism applied to the propensity score have not been subject of a comprehensive study. In our work we try to shed some light on the properties of matching with caliper used to control the differences in the propensity score. Our work is broader than Austin (2009). We consider different distributions of the propensity score and the outcome equation. We also compare the nearest-neighbour matching with and without caliper. It is also worth mentioning that the literature discusses the properties of caliper applied to covariates. The novelty of our approach is that we show properties of caliper imposed on the propensity score. Our main result is that using caliper may lead to biased estimates. Usually, bias is heavy, especially when the outcome value is non-linearly related to the propensity score. Only when either outcome equation is constant or the propensity score distribution is uniform the matching with caliper procedure is able to provide unbiased results. Secondly, the bias becomes smaller as the caliper value increases. Thirdly, to efficiently control for poor matches, the size of the caliper should be within the range recently proposed by Austin (2009).

The article is organised as follows. Section 2 contains short literature review. Section 3 introduces notation, matching estimators and caliper mechanism in detail. In the fourth section we describe Monte Carlo experiment for different distributions of the propensity score and the outcome equations. In the fifth we present our main results. The last section summarises and concludes.

2. LITERATURE REVIEW

Several studies looked into asymptotic properties of matching procedures, to mention Heckman et al. (1998), Hirano, Imbens and Ridder (2003), Abadie and Imbens (2011). All of them show that most matching techniques provide consistent estimates for the average treatment effect; however, only few of them are efficient, such as the class of reweighting estimators. Secondly, pair matching is asymptotically inefficient (Abadie, Imbens, 2006). In a recent article, Abadie and Imbens (2011) showed that simple matching estimators may include bias and that bias does not disappear in large samples. On the other hand, there is only a limited number of

studies dealing with finite sample properties of matching estimators. Here it is worth noting the works by Frölich (2004), Austin (2009) and Busso, DiNardo and McCrary (2009).

Frölich (2004) examined properties of various propensity score matching estimators and showed that one-to-one matching is outperformed by ridge matching. However, the Mean Squared Error (MSE) of ridge matching procedure is lower than that of one-to-one only if the optimal bandwidth is known. Usually, the optimal value of bandwidth is not known *a-priori* and has to be estimated. Austin (2009) also compared several matching techniques in a Monte Carlo study. He concentrated mainly on one-to-one matching. All examined estimators resulted in a similar number of matched pairs and similar balance of variables between treated and untreated samples. Moreover, matching on the propensity score with caliper size not exceeding 0.03 tends to result in estimates with negligible relative bias. Similarly, Busso, DiNardo and McCrary (2009) and Huber, Lechner and Wunch (2013) emphasise the role of trimming to account for common support. Controlling for common support condition effectively improves matching performance regardless of the estimator used.

Deheija and Wahba (1999, 2002) evaluate the performance of propensity score matching methods, including pairwise matching and caliper matching. They find that these simple matching estimators succeed in closely replicating the experimental results. Smith and Todd (2001) reconcile those findings and show that matching estimates cause substantial bias. More recently, Zhao (2004) studied small sample properties of propensity score matching versus covariate matching estimators and those of different matching metrics. He showed that propensity score matching is a good choice when the correlation between covariates and the participation indicator is high. On the other hand, propensity score matching does not perform well in small samples in comparison with other estimators.

Matching estimators frequently suffer from bias. In a seminal work by Rosenbaum and Rubin (1985b) three sources of bias in matching were identified. The first is departure from strong ignorable treatment assignment, which means that assignment to treatment is based on observable pre-treatment variables only. The second one is bias due to incomplete matching and, finally, the third component is due to inexact matching. The bias caused by incomplete matching can be severe and is much worse than the bias due to inexact matching. Assuming that the strong ignorability condition holds, we want to assess empirically which source of matching bias makes larger distortions. Recently, Busso, DiNardo and McCrary (2009) showed that pair matching performs best in terms of bias among all procedures usually applied.

3. CALIPER MATCHING

The main problem in treatment effect literature is the estimation of the average treatment effect on the treated. We follow a standard notation. Let Y_{1i} be an outcome

when individual i receives a treatment and Y_{0i} when he or she does not. The latter situation is called control treatment. Let X be a vector of individual characteristics, $P_i \in \{0,1\}$ be an indicator of treatment status. The treatment effect for an individual i can be written

$$\tau_i = Y_{1i} - Y_{0i}. \tag{1}$$

The problem arise, because for each individual i one can observe either Y_{1i} or Y_{0i}. Hence, estimating the individual treatment effect τ_i is not possible and one has to concentrate on the average treatment effect on the treated (Caliendo, Kopeing, 2008).

The latter is defined as

$$ATT = E\left(Y_1 \mid P = 1, X\right) - E\left(Y_0 \mid P = 1, X\right). \tag{2}$$

A typical matching estimator has the form (Smith, Todd, 2005)

$$\frac{1}{N}\sum_{i=1}^{N}\left[Y_{1i} - E\left(Y_{0i} \mid P_i = 1, X\right)\right], \tag{3}$$

where $E(Y_{0i} \mid P_i = 1, X) = \sum_{j=1}^{N} W(i,j)Y_{0j}$ is an estimator of the counterfactual state, $W(i,j)$ is a matrix of distance between i and j, and N is a number of matched pairs. The fundamental problem of inference is that, for each individual we can observe only one of these potential outcomes, because each unit will receive either treatment or control, not both. The estimation of treatment effects can thus be thought of as a missing data problem (Rubin, 1973), where we are interested in replicating the unobserved potential outcomes.

It is assumed that, conditional on all factors X that influence the potential outcome and the decision to participate, P is independent of Y_0. This assumption has several names in the literature. It is called unconfoundness, conditional independence or overlap, or selection on observables (Imbens, 2004). The counterfactual mean can be identified, provided that the support of X for the treated sample is contained in the support for X in the non-treated one. This property is called common support condition. An additional assumption is the Stable Unit Treatment Value Assumption (Rubin, 1980), which states that the outcomes of one individual are not affected by treatment assignment of any other individual.

The idea of matching is to compute a similarity measure and use the algorithm to match observations from the treatment group with their closest counterpart from the control sample. The aim is to construct an adequate comparison group that replaces missing data and allows us to estimate $E(Y_{0i}|P=1,X)$ without imposing additional *a-priori* assumptions (Blundell and Costa-Dias, 2009). Objects are matched according

to the estimated value of the similarity measure. The straightforward algorithm is to choose for each object in the treatment sample an object with the most proximal value of the similarity measure p_i from the control sample. Usually the propensity score, $p_i = \Pr(P = 1 \mid X_i)$, which is the probability of receiving the treatment by individual i with characteristics X_i, is chosen for that purpose. Let us define a set C_i such that for each unit i only one comparison unit j belongs to C_i:

$$C_i = \left\{ X_{j,j \neq i} \mid j \in \{1 \ldots n\} : \min \left\| p_i - p_j \right\| \right\}, \tag{4}$$

where ||.|| is a metric. In case of the nearest neighbour matching, set C_i can be treated as weighting matrix. The weight matrix $W(i,j)$ is a square matrix with zeros and ones as elements. The value one is assigned to the closest neighbour, and zeros to all remaining units. This type of matching is called one-to-one matching. Each unit from the treatment group is linked with only one element in the control group.

The nearest neighbour matching estimator has good statistical properties if p_i and p_j are defined on a common set. The role of the evaluator is to decide how to treat poorly matched observations (Lee 2005, pp. 89). The total distance, the average distance or the median distance between matched pairs p_i-p_j may be viewed as a measure of matching quality (Rosenbaum, 1985). The lower the measure, the better the fit. For the ideal procedure all quality measures should equal 0. Relying on all matched pairs regardless of matching quality may affect the balance. The balance is a weaker condition than close matching within each pair, and since it is weaker it can often be attained when close matching within pairs is not possible. Rosenbaum and Rubin (1983) showed that balancing two samples on the propensity score is sufficient to equalise covariate distributions. On the other hand, if a large number of poorly matched pairs were left out, the size of the control sample shrinks which means that for certain observations in the treatment group there is no adequate comparison in the control sample. As a result, they are dropped from the analysis. This would help with the balance but at the cost of efficiency, because some information is not used. The evaluator has to choose between the bias due to inexact matching and bias and increased variance due to incomplete matching.

One-to-one or one-to-many matching is characterised by the risk of having poorly matched pairs, that is, pairs distant in terms of the chosen similarity measure. The caliper matching (Cochran, Rubin, 1973) is a variation of the nearest neighbour matching that attempts to avoid poor quality matches by imposing a tolerance of the maximum distance $||p_i$-$p_j||$ allowed. The impact of the caliper may be compared to the focus in a camera. When attention is paid to a specific point, other distant points are not visible. The procedure simply drops objects without a close match in the control group

$$C_i = \left\{ X_{j,j \neq i} \mid j \in \{1 \ldots n\} : \min \left\| p_i - p_j \right\| < \delta \right\}. \tag{5}$$

The set C_i is made of such objects j, that their distance from the nearest match is not greater than δ. That is, a match for person i is selected only if $||p_i-p_j|| < \delta$, where δ is the pre-specified tolerance. Treated persons for whom no matches can be found within caliper are excluded from the analysis, which is one way of imposing a common support condition. Implementation of caliper matching may lead to a smaller bias in regions where similar controls are sparse. An unresolved problem is choosing an *a-priori* reasonable value for tolerance level.

Rosenbaum and Rubin (1985b) discuss the choice of the caliper size, generalizing the results from Table 2.3.1 of Cochran and Rubin (1973). When variance of the linear propensity score in the treatment group is twice as large as that in the control group, a caliper of 0.2 standard deviation removes 98% of the bias in a normally distributed covariate. Rosenbaum and Rubin generally suggest a caliper of 0.25 standard deviation of the linear propensity score. However, in the analysis they considered matching on the Mahalanobis distance not on the propensity score.

Unfortunately, there is no single optimal value for the caliper. The literature suggests small numbers such as 0.005 or 0.001 (see Austin, 2009). The caliper reduces the bias of the average treatment effect estimator at the cost of an increased variance (Heckman et al., 1997). In a special case, when the propensity score distribution is the same in the treatment and the control group, the caliper cuts off the worst matched pairs and lower the bias without a significant increase in the estimator variance. The caliper also lowers the value of matching quality measures. The cost is a lower number of successfully matched pairs. As a consequence, the variance of the average treatment effect may increase. However, this is not a major concern as long as one is interested in a precise estimation of the ATT (Smith, Todd, 2005). On the other hand, Smith and Todd (2005) point out that the potential problem with a caliper is the lack of *a-priori* knowledge about its optimal value. It is common practice to set the value by trial and error.

4. MONTE CARLO STUDY DESIGN

In this section we describe a Monte Carlo simulation conducted to examine the properties of the propensity score matching with caliper and compare with one-to-one matching on the propensity score. Since the propensity score is unknown in general, it is assumed, that it is estimated in a semi-parametric way. In practice, logit, probit or linear probability model is used.

In the Monte Carlo experiment several characteristic of matching can be examined. For instance, different variants of matching procedure, different functional forms of the outcome variable, (in)dependency of the outcome from the propensity score, different distributions of covariates, different sample sizes and different proportion of treated observations to non-treated ones. This complexity makes a general experiment cumbersome. We decided to design the Monte Carlo experiment in such a way that

is as simple as possible on one hand, and as comprehensive as possible on the other. Therefore, we concentrate our attention on the most important features of the matching procedure. We decided not to estimate the propensity score; we instead assume that the propensity score values are drawn from known distributions.

The second pre-set element is the sample size and the proportion of treated to non-treated observations. We decided to work with moderate sample sizes of 500 observations. A number of that range is very common in literature and enables analysis of small sample properties. As shown by Frölich (2004), different proportions of treated and controls may result in different conclusions. We decided to hold this parameter constant and set its value to 1/3. This means that we have twice as many control observations as the treated ones. These proportions are usually found in empirical studies. Usually the control group is larger than the treated group but the difference is not large (see Frölich, 2004).

We considered three different distributions: uniform, normal and Johnson S_B. In the case of two latter distributions, the distribution in the treatment sample is concentrated at the right tail, while in the control sample it is concentrated at the left tail. This setup is used to mimic real differences between the treatment and the control group. The distributions are parameterised and rescaled in such a way that the support is always on the (0,1) interval. They are presented in Figure 1.

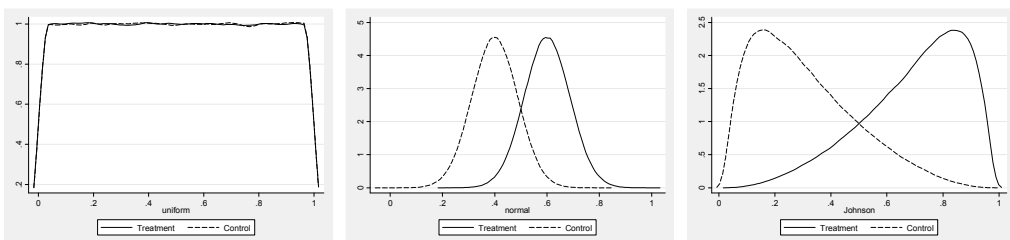


Figure 1. Propensity score distributions
Legend: solid line represent distribution in treated samples, dashed in controls ones.
Source: own computations.

Different distributions are used to replicate the behaviour of real data. The uniform distribution of the propensity score vector, presented on the left panel of Figure 1, is just used as a benchmark. The normal distribution, presented on the middle panel of Figure 1, is a picture of a rather ideal case in which linear combinations of object characteristics follow a normal distribution. In practice, normal distribution is only an approximation and true distribution may possess heavy tails or outliers. Nevertheless, the normal distribution of several characteristics is a common assumption in social sciences. On the right panel the propensity scores follow a Johnson S_B distribution. This is a very flexible distribution, described by four parameters, with a closed analytical form. Depending on the specific parameterisation, Johnson S_B distribution

can be similar to normal distribution, to asymmetric distribution, to distribution with heavy tails, to distribution with probability mass concentrated at the edge of support and to many others. Due to those properties it is frequently used in simulation based studies.

The next element in our numerical experiment is the functional form of the outcome equation. We depart from uniform curve and end up with a highly non-linear outcome. The outcome equations are summarized in Table 1 and presented on Figure 2. The uniform distribution mirrors the ideal case, when the value of treatment is the same for all objects. This distribution will also be used as a benchmark. The linear distribution reflects the situation in which objects that are more likely to take part in a program will benefit more. For instance, this is very common in social support programs. Two other non-linear curves are adapted from Frölich (2004). The m2 curve might represent a situation where the outcome depends discontinuously on an object characteristic strongly related to the propensity score. The m4 curve could be thought of as a reversal of the linear curve. The program pays outmost for those participants that are less likely to participate. Consider, for example, a job training program and education as a key determinant of the propensity score. Usually, well educated persons do not need such programs and are able to find a job without external help. The outcome for the non-treated population is set to be 0 for the reasons of simplicity.

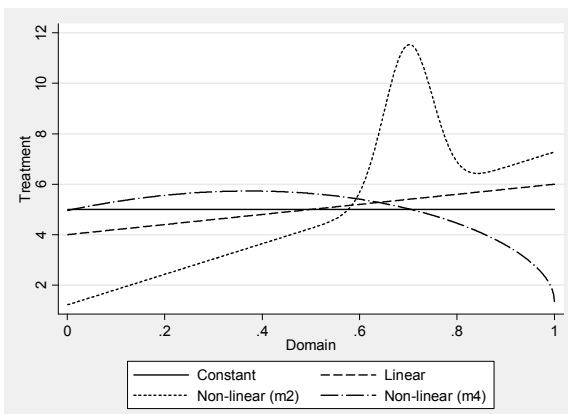


Figure 2. Distribution of the treatment effect

Source: own computations.

In the simulations we use one-to-one procedure and concentrate on caliper size and comparison of matching with and without caliper. The literature suggest rather narrow caliper, but then results are prone to be biased because many potential pairs would be discarded from analysis. The wider caliper allows for greater imbalance at the individual level, however it is easier to achieve balance in the propensity score

distribution between the treated and the control group. Therefore, we considered wide range of caliper sizes from 0.001 to 0.05. Our aim is to check the influence of a particular value of the caliper size on the estimate of ATT.

Table 1.

Outcome equations for the treated sample

Distribution	Outcome equation for the treated group
Constant	y = 5+e, e~U(0, .01)
Linear	y = 4+2*p+e, e~U(0, .01)
Non-linear m2	$y = 0.1+0.5*p+1/2*(\exp(-200*(p-0.7)^2))+e$, e~U(0, .01)
Non-linear m4	$y = 0.2+(1-p)^{0.5}-0.6*(0.9 - p)^2+e$, e~U(0, .01)

Please note that the non-linear curves are adjusted by a linear transformation to have a mean value of 5.

Having assigned specific values to all experiment parameters we are ready to compare the results of the simulation. In each numerical experiment we apply standard one-to-one matching and one-to-one matching with caliper to the same data set. The numerical experiments are designed in such a way that the "true" value of the ATT should be equal to 5 regardless of the distribution of the propensity score, functional form of outcome equation and the estimation technique. The small error term added to the outcome equation is purely random, and hence deviations from 5 no grater than 0.01 are meaningless. As the robustness check we ran simulations with larger random errors of 0.05 and 0.5, but the size of error turned out to have no impact on the final results.

5. EMPIRICAL RESULTS

The main results of our numerical experiment are presented in three separate tables. Each table consist of outcomes for only one distribution of the propensity score and all possible combinations of other parameters. The "1:1" column presents simple one-to-one nearest neighbour matching estimates and the "1:1 caliper" contains the result for matching with caliper. For each caliper size the number in the top row is an estimate of the ATT, while the number in the bottom row is the standard error of that ATT estimate. The results presented in Table 2 serve as kind of a benchmark for further results. They are obtained under the assumption of identical distribution of the propensity score in the treatment and the control group.

When estimated propensity scores have a uniform distribution, identical in the treatment and the control group, neither the shape of the outcome equation nor the size of the caliper have influence on estimates. Both methods, with and without caliper, provide identical results. It is worth emphasising that different caliper values cause a different number of matched pairs (see Table 5).

Table 2.

The ATT estimated with uniform distribution of propensity score

Treatment	constant		linear		m2		m4	
Caliper size	1:1	1:1 caliper	1:1	1:1 caliper	1:1	1:1 caliper	1:1	1:1 caliper
0.001	5.000	5.000	5.000	5.000	5.010	5.010	4.997	4.997
	0.000	0.000	0.045	0.064	0.215	0.307	0.069	0.099
0.005	5.000	5.000	5.000	5.000	5.010	5.011	4.997	4.998
	0.000	0.000	0.045	0.046	0.215	0.219	0.069	0.071
0.010	5.000	5.000	5.000	5.000	5.010	5.010	4.997	4.997
	0.000	0.000	0.045	0.045	0.215	0.215	0.069	0.069
0.020	5.000	5.000	5.000	5.000	5.010	5.010	4.997	4.997
	0.000	0.000	0.045	0.045	0.215	0.215	0.069	0.069
0.025	5.000	5.000	5.000	5.000	5.010	5.010	4.997	4.997
	0.000	0.000	0.045	0.045	0.215	0.215	0.069	0.069
0.050	5.000	5.000	5.000	5.000	5.010	5.010	4.997	4.997
	0.000	0.000	0.045	0.045	0.215	0.215	0.069	0.069

Please note that for each caliper size the number in the top row is an estimate of ATT and the one in the bottom row is its standard error.

Source: own computations.

Table 3.

The ATT estimated with normal distribution of propensity score

Treatment	constant		linear		m2		m4	
Caliper size	1:1	1:1 caliper	1:1	1:1 caliper	1:1	1:1 caliper	1:1	1:1 caliper
0.001	5.000	5.000	5.000	4.846	4.997	3.462	4.999	5.211
	0.001	0.001	0.013	0.016	0.148	0.098	0.021	0.015
0.005	5.000	5.000	5.000	4.897	4.997	3.796	4.999	5.159
	0.001	0.001	0.013	0.012	0.148	0.097	0.021	0.013
0.010	5.000	5.000	5.000	4.917	4.997	3.996	4.999	5.135
	0.001	0.001	0.013	0.012	0.148	0.104	0.021	0.013
0.020	5.000	5.000	5.000	4.934	4.997	4.224	4.999	5.112
	0.001	0.001	0.013	0.011	0.148	0.113	0.021	0.013
0.025	5.000	5.000	5.000	4.940	4.997	4.307	4.999	5.104
	0.001	0.001	0.013	0.011	0.148	0.118	0.021	0.013
0.050	5.000	5.000	5.000	4.959	4.997	4.611	4.999	5.074
	0.001	0.001	0.013	0.012	0.148	0.137	0.021	0.014

Please note that for each caliper size the number in the top row is an estimate of ATT and the one in the bottom row is its standard error.

Source: own computations.

In case of a normal distribution of the propensity score the results for one-to-one matching are similar to those in Table 2, despite the fact that the propensity score distribution in the treatment and the control group is different. Unfortunately, the caliper mechanism seems to cause bias to the results. The caliper estimates are unbiased for constant treatment. The effect of bias is moderate in case of linear treatment and non-linear m4 treatment. For non-linear m2 treatment equation the bias is heavy. In all cases the bias becomes smaller as the caliper size rises.

Those results indicate that the caliper mechanism can potentially distort the estimation results.

The last set of estimates present the results for a Johnson S_B distribution of the propensity score (Table 4.). In this simulation the pre-matching differences between the treated and the control group are the greatest. Despite that, one-to-one matching is able to provide unbiased estimates for all but one distribution of the outcome. The estimates for a non-linear m2 curve are biased and the bias significantly exceeds randomness of simulation. The results seem to indicate that matching with caliper is a much worse choice. Estimates are biased for all non-constant outcomes. For a linear and a non-linear m4 specification the bias of caliper matching is relatively large. In case of a non-linear m2 outcome the bias is larger than that of one-to-one matching.

Table 4.

ATT estimated with Johnson S_B distribution of propensity score

Treatment	constant		linear		m2		m4	
Caliper size	1:1	1:1 caliper	1:1	1:1 caliper	1:1	1:1 caliper	1:1	1:1 caliper
0.001	5.000	5.000	5.001	4.694	5.089	4.197	5.003	5.686
	0.000	0.000	0.026	0.052	0.135	0.322	0.068	0.073
0.005	5.000	5.000	5.001	4.809	5.089	4.796	5.003	5.523
	0.001	0.001	0.026	0.032	0.135	0.213	0.068	0.055
0.010	5.000	5.000	5.001	4.863	5.089	5.018	5.003	5.418
	0.001	0.001	0.026	0.029	0.135	0.189	0.068	0.054
0.020	5.000	5.000	5.001	4.906	5.089	5.099	5.003	5.314
	0.001	0.001	0.026	0.027	0.135	0.170	0.068	0.054
0.025	5.000	5.000	5.001	4.918	5.089	5.103	5.003	5.281
	0.001	0.001	0.026	0.027	0.135	0.165	0.068	0.055
0.050	5.000	5.000	5.001	4.952	5.089	5.094	5.003	5.180
	0.001	0.001	0.026	0.026	0.135	0.151	0.068	0.058

Please note that for each caliper size the number in the top row is an estimate of ATT and the one in the bottom row is its standard error.

Source: own computations.

Those results indicate that controlling for strict similarity of the estimated propensity score is not sufficient to obtain unbiased estimates of the ATT. This in turn implies that, in small samples, bias arising from inexact matching is relatively low in comparison with the one caused by incomplete matching, This is not an extraordinary finding, since in small samples the number of successfully matched pairs is low and each matched pair approximately accounts for 1% of total pairs. If one removes 20% or even 50% of total initial pairs this may cause a spectacular bias.

Table 5 illustrates the problem of diminishing matched pairs. As the caliper size shrinks, the number of successfully matched pairs decreases dramatically if the propensity scores are differently distributed in the treatment and in the control group. With the caliper value of 0.005, which is the one most frequently suggested in the literature, one would lose over 40% of matched pairs in case of a normal distribution of the propensity score, and nearly 45% in case of a Johnson S_B specification. The bias of the estimate is the result of a non-symmetric distribution of the outcome.

Table 5.

Number of successfully matched pairs

	Propensity score distribution					
	uniform		normal		Johnson S_B	
Caliper size	1 to 1 matching	caliper matching	1 to 1 matching	caliper matching	1 to 1 matching	caliper matching
0.001	165	81	165	52	165	83
0.005	165	159	165	96	165	93
0.010	165	165	165	112	165	115
0.020	165	165	165	126	165	132
0.025	165	165	165	130	165	136
0.050	165	165	165	140	165	149

Source: own computations.

To compare efficiency of both methods we decided to compute the RMSE statistics, presented in Table 6. With uniform distribution of the propensity score the RMSE for caliper matching is larger than for no caliper for all considered outcome equations. Results of simulations are similar for a normal and a Johnson S_B distribution of the propensity score. The caliper mechanism provides better results than matching without caliper for constant treatment. For all other functional forms of treatment equation the results showed that simple matching provides more precise estimates.

We derive two important implications from the above results. First, we confirm the theoretical results which implicitly assume constant character of treatment. Secondly, and more importantly, the caliper mechanism induces significant rise in the bias and the variance of the estimates in case of a non-uniform treatment. In empirical practice

it is very unlikely that the impact of treatment can be treated as uniform. Therefore, the usage of caliper introduces variance and this variance is larger than the reduction of variance due to lowering the bias.

Table 6.

Root Mean Squared Error

Treatment	constant		linear		m2		m4	
Distribution	no caliper	standard caliper	no caliper	standard caliper	no caliper	standard caliper	no caliper	standard caliper
uniform	0.000376	0.000378	0.045101	0.046571	0.216150	0.223129	0.070145	0.072525
normal	0.001139	0.000659	0.013216	0.155195	0.146735	1.223241	0.020911	0.160712
Johnson	0.000906	0.000571	0.026516	0.311774	0.163154	0.334864	0.068231	0.531482

RMSE computed for caliper size of 0.005.

Source: own computations.

6. CONCLUSIONS

In this article we studied properties of matching and matching with caliper in small samples. Despite that not much is known, both methods are widely used in applied evaluation research and it is important to understand their properties. We focused our attention on the properties of the ATT estimator. There are relatively few studies on the small-sample properties of different matching estimators. We tried to bridge the gap and shed some light on the problem.

To achieve that, we have used distributions that are either assumed in theoretical papers or employed in similar simulations. We confirmed the theoretical results and showed that simple one-to-one matching estimators are unbiased in most cases. At the same time the caliper method which is theoretically designed to control for the differences in distributions of the propensity score may introduce a substantial bias to the ATT estimator in small samples.

It turned out that in small samples the bias due to inexact matching is relatively small in comparison with that of incomplete matching. There are several reasons for this. First of all, there are only few unmatched observations. Secondly, if the covariates have the same distribution among matched and unmatched units the bias is limited. The third possibility is a constant value of treatment. Our empirical results suggest that even if the treatment is not constant, the value of bias is not large. On the other hand, the bias due to incomplete matching turned out to be substantial in our simulations, up to 15% for very conservative caliper size, and about 10% in case of the most popular 0.005 caliper.

Another practical problem with using a caliper is the loss of a significant number of possible matched pairs. This is of particularly great importance when the pool

of possible matched pairs is limited. A decrease in the number of matches caused by caliper is the primary reason why the caliper matching becomes incomplete. Moreover, the results show that this lack of completeness causes a significant bias to the ATT estimates.

There are several limitations to our simulations. First of all, we assumed that we work on propensity scores, not covariates. Our results are valid as long as distributions used in a simulation to mimic behaviour of the real data are close to empirical realisations. To achieve that, we have used distributions that are either assumed in empirical works or used in similar studies. To provide robustness of our results we alter the parametrisation of the distributions, and the results seem to be robust. Secondly, we concentrate on small sample behaviour of the ATT estimator. It is worth noting that 500 observations used in simulations give a maximum of 165 matched pairs. This number is rather low. Therefore, we replicate the part of the experiment for Johnson S_B distribution and a sample size of 1,500 (about 500 matched pairs) and the general picture does not change, however biases are much lower, about a half of those presented in Table 4. The replication of full experiment is hardly possible due to computational complexity. Nevertheless, the robustness is in our opinion confirmed with these results.

University of Warsaw

REFERENCES

[1] Abadie A., Imbens G., (2006), Large Sample Properties of Matching Estimators for Average Treatment Effects, *Econometrica*, 74 (1), 235–267.

[2] Abadie A. Imbens G., (2011), Bias-Corrected Estimates for Average Treatment Effects, *Journal of Business and Economic Statistics*, 29, 1–11.

[3] Austin P. (2009), Some Methods of Propensity Score Matching Had Superior Performance to Others: Result of an Empirical Investigation and Monte Carlo Simulations, *Biometrical Journal*, 5, 171–184.

[4] Blundell R., Costa-Diás M., (2000), Evaluation Methods for Non-Experimental Data, *Fiscal Studies*, 21 (4), 427–468.

[5] Busso M., DiNardo J., McCrary J., (2009), New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators, *IZA Discussion Paper*, 3998.

[6] Cochrane W., Rubin D., (1973), Controling Bias in Observational Studies. A Review, *Sankhya*, 35, 417–466.

[7] Dehejia R., Wahba S., (1999), Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Program, *Journal of American Statistical Association*, 94, 1053–1062.

[8] Dehejia R., Wahba S., (2002), Propensity Score Matching Methods for Nonexperimental Causal Studies, *Journal of the American Statistical Association*, 84, 151–161.

[9] Frölich (2004), Finite Sample Properties of Propensity-Score Matching and Weighting Estimators, *The Review of Economics and Statistics*, 86 (1), 77–90.

[10] Heckman J., Ichimura H., Smith J., Todd P., (1998), Characterizing Selection Bias Using Experimental Data, *Econometrica*, 66 (5), 1017–1098.
[11] Heckman J., Ichimura H., Todd P., (1997), Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *The Review of Economic Studies*, 64 (4), 605–654.
[12] Hirano K., Imbens G., Ridder G., (2003), Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score, *Econometrica*, 71 (4), 1161–1189.
[13] Huber D., Lechner M., Wunch C., (2013), The Performance of Estimators Based on the Propensity Score, *Journal of Econometrics*, 175 (1), 1–21.
[14] Imbens G., (2004), Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review, *Review of Economics and Statistics*, 86 (1), 4–29.
[15] Lee M-J., (2005) *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford University Press.
[16] Rosenbaum P., Rubin D., (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, 70 (1), 41–55.
[17] Rosenbaum P., Rubin D., (1985a), Constructing Control Group Using Multivariate Matched Sampling Methods That Incorporate Propensity Score, *The American Statistician*, 39 (1), 33–38.
[18] Rosenbaum P., Rubin D., (1985b), Bias Due to Incomplete Matching, *Biometrics*, 41 (1), 103–116.
[19] Rubin D., (1973), Matching to Remove Bias in Observational Studies, *Biometrics*, 29, 159–183.
[20] Rubin D., (1980), Bias Reduction Using Mahalanobis Metric Matching, *Biometrics*, 36 (2), 293–298.
[21] Smith J., Todd P., (2001), Reconciling Conflicting Evidence on the Performance of Propensity--Score Matching Methods, *The American Economic Review*, 91 (2), 112–118.
[22] Smith J., Todd P., (2005), Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?, *Journal of Econometrics*, 125, 305–353.
[23] Zhao Z., (2004), Using Matching to Estimate Treatment Effects: Data Requirements, Matching Methods, and Monte Carlo Evidence, *The Review of Economics and Statistics*, 86 (1), 91–107.

WŁASNOŚCI MAŁOPRÓBKOWE ESTYMACJI PRZEZ DOPASOWANIE

Streszczenie

Mechanizm obcięcia (suwmiarki) jest szeroko stosowanym narzędziem zabezpieczającym przed słabo dopasowanymi połączeniami. W literaturze opisywane są własności asymptotyczne estymatorów z obcięciem. W artykule opisany jest eksperyment symulacji numerycznej badający własności tych estymatorów w małych i przeciętnie liczebnych próbach. Pokazujemy, że mechanizm obcięcia (suwmiarki) powoduje znaczne obciążenie estymatora przeciętnego efektu oddziaływania wobec jednostek poddanych oddziaływaniu (*ang*. ATT) i wzrost wartości jego wariancji w porównaniu ze standardowym estymatorem 1:1.

Słowa kluczowe: łączenie według prawdopodobieństwa, obcięcie (suwmiarka), eksperyment Monte Carlo, własności małej próby

SMALL SAMPLE PROPERTIES OF MATCHING WITH CALIPER

A b s t r a c t

A caliper mechanism is a common tool used to prevent from inexact matches. The existing literature discusses asymptotic properties of matching with caliper. In this simulation study we investigate properties in small and medium sized samples. We show that caliper causes a significant bias of the ATT estimator and raises its variance in comparison to one-to-one matching.

Keywords: propensity score matching, caliper, Monte Carlo experiment, finite sample properties

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

MARCIN ANHOLCER, HELENA GASPARS-WIELOCH

ACCURACY OF THE KAUFMANN AND DESBAZEILLE ALGORITHM FOR TIME-COST TRADE-OFF PROJECT PROBLEMS

1. INTRODUCTION

The Kaufmann and Desbazeille algorithm is a procedure used in the time-cost trade-off project analysis (TCTP-analysis). The method in question is commonly known by project management trainers, academic teachers and students. The algorithm is presented, among other places, in Bladowski (1970), Gaspars (2006a), Gaspars (2006b), Hendrickson (1989), Idźkiewicz (1967), Kukuła et al. (1996), Muller (1965), Muller (1964), Siudak (1998) and Trocki et al. (2003). Authors use different names for this method (e.g. CPM-COST analysis, time-cost analysis, network compression algorithm, MCX – Minimum Cost Expediting). That is why Kaufmann and Desbazeille (1964) are little-known surnames in TCTP analysis. Nevertheless, their book is the oldest one containing a description of the algorithm under consideration. Therefore, the procedure analyzed in this article, is conventionally called the Kaufmann and Desbazeille algorithm. For convenience, we also use of the abbreviation KDA.

Many people claim that this method is an exact algorithm (Bladowski, 1970; Bozarth, Handfield, 2005; Fusek et al., 1967; Giard, 1991; Idźkiewicz, 1967; Kopańska-Bródka, 1998; Kukuła et al., 1996; Muller, 1965; Trocki et al., 2003; Waters, 1998), i.e. an algorithm which always indicates the cheapest way of project compression. The KDA is however only a heuristic procedure. Examples of problems for which the method leads to quasi-optimal solutions are given in Gaspars (2006a) and Gaspars (2006b).

In Anholcer, Gaspars-Wieloch (2011) we have discussed and proved the special case for which the Kaufmann and Desbazeille algorithm gives the worst results. In this paper we are going to calculate the average accuracy of the KDA for several test problems similar to the worst case and for some randomly generated problems, as well.

2. TIME-COST TRADE-OFF PROJECT PROBLEMS

In TCTP problems the trade-off occurs between the project completion time (T) and the amount of non-renewable resources, i.e. money, which constitute the total cost of the project (TC).

The total cost (*TC*) consists of direct (*DC*) and indirect costs (*IC*). The first category concerns costs related directly to the completion of activities (e.g. labor, raw materials) and their compression. Activity durations are bounded from below (crash duration) and from above (normal duration). The shortening of a given task requires the increase of the amount of resources that are used to accomplish this activity. Indirect costs, like e.g. penalty, insurance and taxes, are assigned to the project as a whole. The penalty costs are imposed when the completion time has been delayed. In general, when the completion time increases, indirect costs increase and direct costs decrease (see Figure 1).

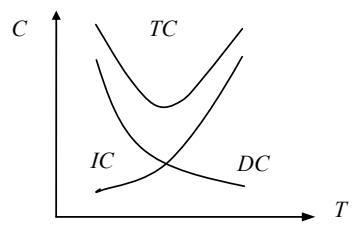


Figure 1. Time-cost curves

The main goals considered in TCTP analysis are as follows (Skutella, 1998):

1) minimizing time-dependent project costs within a specified project deadline or target time T^d (Deadline Problem),
$$C(X) \to \min, \tag{1}$$
$$T(X) \leq T^d, \tag{2}$$

2) minimizing project completion time within a specified budget C^b (Budget Problem),
$$T(X) \to \min, \tag{3}$$
$$C(X) \leq C^b, \tag{4}$$

where X signifies the vector of activity durations.

3. EXISTING ALGORITHMS FOR THE TCTP PROBLEM

A detailed survey and analysis of optimization methods applicable to the TCTP problems was presented by one of us in Gaspars-Wieloch (2008) and Gaspars-Wieloch (2009). In addition, we both showed a comparative breakdown of these algorithms in Anholcer, Gaspars-Wieloch (2011). Here we just recapitulate the most important parts of the breakdown mentioned above (t^c_i – the shortest possible time for the i-th activity, i.e. crash time; t^n_i – the normal time for the i-th activity), see Table 1.

Table 1.

Comparative breakdown of existing project time-cost trade-off algorithms

Criterion				Authors of procedures
1.	Type of optimization problems solved	$C(X) \rightarrow \min$, $T(X) \leq T^d$		1. Berman 2. Bladowski 3. De, Dunne, Gosh and Wells 4.Falk and Horowitz 5. Fondhal 6. Fulkerson 7. Gedymin 8. Goyal 9. Hindelang and Muth 10. Kaufmann and Desbazeille 11. Kelley 12. Liu, Burns and Feng 13. Moder and Phillips 14. Moussourakis and Haksever 15. Panagiotakopoulos 16. Phillips and Dessouky 17. Prager 18. Siemens
		$T(X) \rightarrow \min$, $C(X) \leq C^d$		1. Bladowski 2. De, Dunne, Gosh and Wells 3. Fulkerson 4. Gedymin 5. Kaufmann and Desbazeille 6. Kelley 7. Liu, Burns and Feng 8. Moder and Phillips 9. Moussourakis and Haksever 10. Phillips and Dessouky 11. Prager
		$C(T^{ad}) \rightarrow \min$ *		1. Bladowski 2. Crowston and Thompson 3. De, Dunne, Gosh and Wells 4. Fulkerson 5. Gedymin 6. Hindelang and Muth 7. Kaufmann and Desbazeille 8. Kelley 9. Liu Burns and Feng 10. Moder and Phillips 11. Moussourakis and Haksever 12. Phillips and Dessouky 13. Prager
2.	Type of costs considered	Shortening costs		1. Bladowski 2. Fulkerson 3. Gedymin 4. Goyal 5. Kaufmann and Desbazeille 6. Kelley 7. Panagiotakopoulos 8. Phillips and Dessouky 9. Prager 10. Siemens
		Shortening costs and other direct project costs		1. Berman 2. Bladowski 3. De, Dunne, Gosh and Wells 4. Falk and Horowitz 5. Fondhal 6. Kaufmann and Desbazeille 7. Liu, Burns and Feng 8. Moussourakis and Haksever
		All costs (direct and indirect)		1. Crowston and Thompson 2. Hindelang and Muth 3. Moder and Phillips
3.	Growth speed of direct costs	Constant (linear time-cost curve)		1. Fulkerson 2. Goyal 3. Kelley 4. Phillips and Dessouky 5. Prager 6. Siemens
		Increasing (convex curve)	Linear approximation	1. Kelley 2. Goyal 3. Liu, Burns and Feng 4. Phillips and Dessouky 5. Prager 6. Siemens
			No approximation	1. Berman
		Decreasing (concave curve)	Linear approximation	1. Falk and Horowitz 2. Gedymin
			No approximation	–
		Various (ex. Concave-convex curve)	Linear approximation	1. Falk and Horowitz 2. Gedymin 3. Moussourakis and Haksever
			No approximation	1. Bladowski 2. Crowston and Thompson 3. De, Dunne, Gosh and Wells 4. Fondhal 5. Hindelang and Muth 6. Kaufmann and Desbazeille 7. Moder and Phillips 8. Panagiotakopoulos

Table 1.

Criterion			Authors of procedures
4.	Feasible activity durations	Any real number from $<t^c_i, t^n_i>$	1. Berman 2. Falk and Horowitz 3. Fondhal 4. Fulkerson 5. Gedymin 6. Goyal 7. Kelley 8. Phillips and Dessouky 9. Prager 10. Siemens
		Any integer from $<t^c_i, t^n_i>$	1. Bladowski 2. Kaufmann and Desbazeille
		Discrete options from $<t^c_i, t^n_i>$	1. Crowston and Thompson 2. De, Dunne, Gosh and Wells 3. Hindelang and Muth 4. Liu, Burns and Feng 5. Moder and Phillips 6. Moussourakis and Haksever 7. Panagiotakopoulos
5.	Accuracy of solutions obtained	Optimal solutions	1. Berman 2. De, Dunne, Gosh and Wells 3. Falk and Horowitz 4. Fulkerson 5. Gedymin 6. Kelley 7. Liu, Burns and Feng 8. Moussourakis and Haksever 9. Phillips and Dessouky 10. Prager
		Optimal or suboptimal solutions	1. Bladowski 2. Crowston and Thompson 3. Fondhal 4. Goyal 5. Hindelang and Muth 6. Kaufmann and Desbazeille 7. Moder and Phillips 8. Panagiotakopoulos 9. Siemens

* T^{ad} denotes an advisable, but not mandatory, project completion time. When the advisable time is exceeded, than the total project cost must include a penalty. When the project is finished before the advisable time, the total cost is reduced by a bonus.

Source: Anholcer, Gaspars-Wieloch (2011), Gaspars-Wieloch (2008), Gaspars-Wieloch (2009).

Exact algorithms (i.e. those which guarantee finding an optimal solution) applied in discrete problems (*DTCTP* – Discrete Time-Cost Trade-off Problems) are characterized by an exponential complexity and are NP-hard. Procedures designed for continuous cases are solvable in polynomial time (see De et al., 1995; De et al., 1997; Gaspars-Wieloch, 2008; Panagiotakopoulos, 1977; Siudak, 1998; Skutella, 1998).

4. THE PROJECT NETWORK – NOTATION

In the paper we use the AOA (activities on arcs) network representation of the project, similarly as in Anholcer, Gaspars-Wieloch (2011). To be more exact, the set E of arcs consists of m arcs $e_1, \ldots, e_m$, while the set of nodes V consists of n nodes $v_1, \ldots, v_n$. Each arc e_i, $i = 1, \ldots, m$ is labeled by some positive natural number t_i, i.e. the duration of the respective activity. For convenience, we often index the arc by e_{jk}, where $j = 1, \ldots, n-1$ and $k = 2, \ldots, n$ are the indices of the starting and end node of the arc, respectively.

In addition, for each arc we define a non-decreasing sequence $C^{(i)} = \left(c^{(i)}_{k_i}\right)_{k_i=1}^{K_i}$ of real numbers representing the shortening cost, where $c^{(i)}_{k_i}$ is the cost of reducing the

duration of the i-th activity by the k_i-th unit and $K_i \leq t_i$ (one may shorten an activity at most $(t^n_i - t^c_i)$ times).

The earliest time of the event j (the earliest time at which node j can be reached such that all its preceding activities have been finished) is denoted by t^I_j (t^I_n being equal to the minimum completion time of the project, i.e. T^*). The latest time of the event j (the latest time that node j can be left such that it is still possible to finish the overall project in the minimum completion time) is denoted by t^{II}_j.

5. THE KAUFMANN AND DESBAZEILLE ALGORITHM – DESCRIPTION AND IMPLEMENTATION

The Kaufmann and Desbazeille algorithm is one of the oldest procedures applied in TCTP analysis. As one can notice in the Table 1, the KDA allows to solve the problem (1)-(2) or (3)-(4). This procedure focuses on the minimization of direct costs (see Section 2). Theoretically, the algorithm may be used in time-cost trade-off problems with any types of time-cost curves for project activities. However, in practice it is applied only when the unit shortening costs are non-decreasing, because in other cases the solutions obtained are extremely bad (i.e. far from the optimal ones). Most of the existing exact algorithms for TCTP are designed just for linear or convex time-cost curves.

Kaufmann and Desbazeille assume that the parameters representing the normal activity durations (t_i^n), the project target time (T^d) and the difference (t_i^n-t_i^c) are integers. They take into consideration only integer realizations of the project time even if, for a given C^b (see the Budget Problem (3)-(4)), the optimal solution requires non-integer times for some tasks (see Table 1, point 4).

The original version of the KDA consists in iteratively shortening each critical path in the network by exactly one time unit. The target is to find the cheapest way of compressing the project time by one unit. We stop when constraints (1)-(2) or (3)-(4) are satisfied. In order to apply this procedure, the user should know:

– the structure of the project network,
– the normal and the crash duration of the activities within the project,
– the unit shortening cost for each task,
– the deadline or budgetary constraint.

Let us recall the details of the KDA. Two steps are performed in every iteration: one is to implement the CPM method for the actual durations of activities and the second one is the shortening of the project by one time unit. The algorithm stops when the desired time (T^d) has been reached.

To compress the project duration, it is necessary to reduce each critical path by one time unit. It is not necessary to shorten one activity from each critical path as the critical paths not always are disjoint. In order to reduce the project duration, all the cuts of the critical sub-network are considered and the cheapest one is chosen. Then,

all the activities belonging to the minimal cut are shortened by one time unit. In the original KDA the cut P is defined as the set of arcs such that:

1. After removing all the arcs belonging to P, the project network is no longer connected.
2. The set of nodes V splits into two disjoint subsets: V_1 containing the starting event and V_2 containing the end event of the project such that the starting nodes of all the arcs from P belong to V_1, while the end nodes belong to V_2.

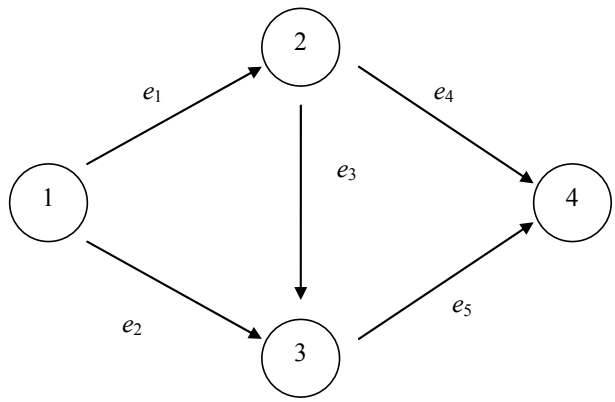


Figure 2. Sample network

In the articles Gaspars (2006a) and Gaspars (2006b) one of us came to the conclusion that the main (but not the only one!) factor affecting the accuracy of the KDA is the way in which variants for compressing the completion time are found. In the original version of the KDA the best set of activities to shorten is chosen from a list of sets containing exactly one activity on each critical path. This approach exposes us to overlooking the cheapest way of accelerating the project by one time unit. In Gaspars (2006a) and Gaspars (2006b) it has been emphasized that the easiest, but not sufficient, way of improving the results, i.e. reducing the compression costs, is to find the cheapest solution among sets containing *at least* (but not exactly) one activity on each critical path. This modification is so natural and obvious that in the next section we will analyze the accuracy of this modified version of the KDA. That means that in order to increase the exactitude of the algorithm it is desirable to consider a more general concept of a cut, as Ford and Fulkerson do (see e.g. Cormen et al., 2001; Ford, Fulkerson, 1962; Gedymin, 1974; Korzan, 1978). To be more specific, they only use the second point of the last definition. This slight modification allows us to use the FFEK[1] algorithm (see e.g. Edmonds, Karp, 1972) for finding the minimal cut instead of an exhaustive search over all the cuts, which makes the algorithm much faster. To

[1] Edmonds – Karp algorithm, one of the adaptations of the Ford – Fulkerson algorithm.

illustrate the difference between these two definitions of the cuts see Figure 2 below. In the case of the original KDA the cuts allowed are $\{e_1, e_2\}$, $\{e_2, e_3, e_4\}$ and $\{e_4, e_5\}$, while in the case of the Ford – Fulkerson algorithm there is one more cut allowed, namely $\{e_1, e_5\}$.

Finally, the version of the Kaufmann – Desbazeille algorithm tested in the next section can be defined as follows:

1. (Initialization) Set the total shortening cost $TC := 0$ and go to step 2.
2. (CPM) Perform the CPM analysis of the project network. If $t_n^I \leq T^d$ then STOP. The current solution is the optimal one[2]. Otherwise go to step 3.
3. (Time reduction / FFEK) Define the following maximum flow problem. Considered arcs in the network are all the critical arcs in the project network. Moreover the capacity of each arc e_i belonging to this network is equal to the first element of the sequence $C^{(i)}$ (if this sequence has at least one element and ∞ otherwise). Using the FFEK method one finds the minimal cut. If the value CV of the minimal cut MC is equal to ∞, then STOP. The deadline problem cannot be solved[3]. Otherwise, go to step 4.
4. (Update) Set $TC := TC + CV$. For each arc $e_i \in MC$ set $t_i := t_i - 1$ and remove the first element of $C^{(i)}$. Go back to step 2.

Note that each time step 4 is performed, the total completion time of the overall project decreases by one. Thus the version of the KDA defined above always stops after a finite number of steps: it either finds the optimal solution or reports the problem to be inconsistent.

6. GOAL OF COMPUTATIONAL EXPERIMENTS

Kaufmann and Desbazeille were only interested in integer realizations of the project time. Therefore, the empirical research will just be related to the deadline problem (see problem 1-2) as the discrete solutions obtained simultaneously constitute solutions for different values of the budgetary parameter in the budget problem. Solutions generated for one of the chosen problem (deadline or budget) suffice to establish the whole time-cost project curve DC (see Figure 1).

In our article we calculate the mean deviation between KD solutions and optimal results. The measure of accuracy (A) is a relative difference between the total project compression costs obtained by the KDA, i.e. C^{KDA}, and the minimum project compression costs C^{min} (the average is calculated over P, i.e. all the test problems, see Equation 5):

---

[2] It is optimal in the sense of the KDA, while it does not have to be the global minimum of the defined deadline problem.

[3] It cannot be solved using the KDA method. It can be easily proved then that the deadline problem is contradictory (inconsistent).

$$A = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{C_p^{KDA} - C_p^{\min}}{C_p^{\min}}\right). \tag{5}$$

The values C_p^{min} constitute the optimal solutions of the following deadline problem:

$$\sum_{i=1}^{m}\sum_{k=1}^{K(i)} c_{ik} y_{ik} \to \min, \tag{6}$$

$$-x_{p(i)} + x_{q(i)} + \sum_{k=1}^{K(i)} y_{ik} - z_i = t_i^n, i = 1,...,m, \tag{7}$$

$$x_1 = 0, x_j \geq 0, j = 1,...,n, x_n \leq T^d, \tag{8}$$

$$0 \leq y_{ik} \leq 1, i = 1,...,m, k = 1,...,K(i), \tag{9}$$

$$z_i \geq 0, i = 1,...,m. \tag{10}$$

with parameters:
m – number of activities (arcs),
n – number of events (nodes),
t_i^n, t_i^c – the normal and crash time of activity e_i, $1 \leq i \leq m$,
$K(i)$ – maximum number of units that activity e_i can be shortened by, $K(i) = t_i^n - t_i^c$,
c_{ik} – cost of shortening activity e_i by k^{th} unit, $c_{ik} \leq c_{i,k+1}$, $1 \leq k < K(i)$,
$p(i)$, $q(i)$ – indices of the starting and end node of arc e_i, $1 \leq p(i) < q(i) \leq n$,
T^d – the desired project completion time,

and variables:
x_j – time of event v_j, $1 \leq j \leq n$,
y_{ik} – binary variable equal to 1 when activity e_i is shortened by k^{th} unit and 0 otherwise,
z_i – final time reserve for activity e_i, $1 \leq i \leq m$.
Each optimal value of the variable x_j always belongs to the interval $[t_j^I, t_j^{II}]$.

In connection with the assumptions made by Kaufmann and Desbazeille (the time parameters are integer) and the fact that we only analyze the deadline problem, the variables in the model (6)-(10) are always integer although there are no additional constraints requiring integer solutions (see Schrijver, 2003). In our experiments we intend to check the accuracy of the KDA for problems similar to the worst case

presented and proved in Anholcer, Gaspars-Wieloch (2011) and for totally random problems.

7. DESCRIPTION OF THE TEST PROBLEMS

Let us recall the details of the case for which the accuracy of the KDA is the worst (see Anholcer, Gaspars-Wieloch, 2011):

1) the network has the structure given in Figure 3 (i.e. the set of arcs consists of all the pairs of nodes of the form (v_j, v_{j+2}), $j = 1, \ldots, n\text{-}2$, and all the pairs of nodes of the form (v_j, v_{j+1}), $j = 1, \ldots, n\text{-}1$) and the number of nodes (n) is even,
2) the normal completion times of activities $<k,k+1>$ equal $t^n_{k,k+1}$ and the normal completion times of activities $<k,k+2>$ equal $2 \cdot t^n_{k,k+1} - 1$ (see Figure 3),
3) the unit shortening costs of activities $<2k, 2k + 1>$, $<2k - 1, 2k>$ and $<k, k + 2>$ are respectively equal to a, b and c where $b > a$ and $c > b\left(\frac{n}{2} - 1\right)$ (see Figure 3),
4) the desired project completion time T^d is equal to $T^n - \left(\frac{n}{2} - 1\right) - 1$, where T^n is the normal project time (i.e. $T^n = (n-1) \cdot t^n_{k,k+1}$)) and $\left(\frac{n}{2} - 1\right)$ is the number of all activities $<2k, 2k + 1>$. This means that $T^d = T^n - \frac{n}{2} = (n-1) \cdot t^n_{k,k+1} - \frac{n}{2}$.

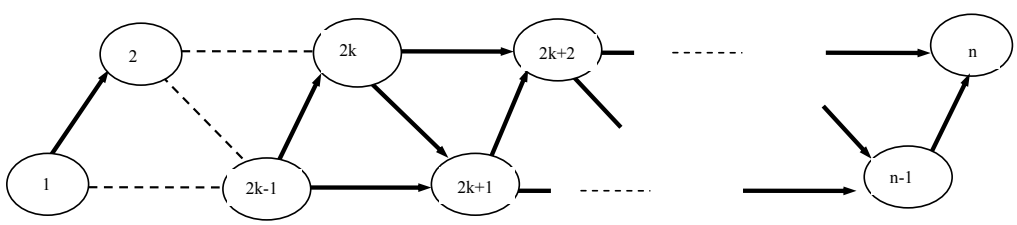


Figure 3. Network type: deterministic

According to the proof given in Anholcer, Gaspars-Wieloch (2011), in such a situation the difference between the total shortening cost given by KDA and the optimal one amounts to $C^{KDA} - C^{\min} = a\left(\frac{n}{2} - 1\right)$. This implies that the gap increases arbitrarily if either n, a or both increase. In particular we have

$$\lim_{n \to \infty}\left(C^{KDA} - C^{\min}\right) = \infty .$$

We use three types of test problems in our experiments. The structure of the first type network is deterministic and is connected to the worst case. Two others are random.

1. Network type 1 (deterministic): The set of nodes consists of n nodes, where n is a parameter. The set of arcs (activities) consists of all the couples of nodes of the form (v_j, v_{j+2}), $j = 1, \ldots, n\text{-}2$, and all the couples of nodes of the form (v_j, v_{j+1}), $j = 1, \ldots, n\text{-}1$ (see Figure 3).
2. Network type 2 (globally random): The set of nodes consists of n nodes, where n is a parameter. There is also one more parameter – D, being the maximum out-degree[4]. The process of generating the arcs consists of two steps. In the first step, for each node v_j, $j = 1, \ldots, n\text{-}D$, we choose uniformly at random D successors from the set $\{v_{j+1}, \ldots, v_n\}$. Each remaining node, i.e. node v_j, $j = n\text{-}D+1, \ldots, n\text{-}1$, is linked with all the nodes from the set $\{v_{j+1}, \ldots, v_n\}$. One may observe that such a way of the network generation does guarantee neither the connectivity of the network nor the uniqueness of the starting event. Thus in the second step all the nodes from the set $\{v_2, \ldots, v_n\}$ are examined. If some of them, say v_j, does not have any predecessor, then an additional arc is added in the following way. A vertex, say v_k, is chosen uniformly at random from the set $\{v_1, \ldots, v_{j-1}\}$, and the new arc is formed as (v_k, v_j).
3. Network type 3 (locally random): The way of generating the arcs is almost the same as in the previous case. The only difference is that in both steps the probabilities of choosing the neighbors are not equal, but the smaller is the difference between the nodes indices, the higher is the probability of joining them with an arc. Namely, the probability of creating an arc joining the pair of nodes (v_j, v_{j+k}) is k times lower than the probability of the creation of the arc between the nodes v_j and v_{j+1}.

In the deterministic case the integer times are chosen uniformly at random from the given range, where the minimal and maximal values are parameters. The ranges are defined separately for the arcs of the type (v_j, v_{j+1}) and the arcs of the type (v_j, v_{j+2}). In the case of random networks there is one basic range defined for all the arcs of the type (v_j, v_{j+1}). For the arcs joining the vertices (v_j, v_{j+k}), $k \geq 1$, the length of the range is calculated as the floor[5] from the product of the basic time and one of the expressions: k (linear growth), $\ln k + 1$ (logarithmic growth), $k^{1/2}$ (square root growth) or $k^{3/2}$ (power growth).

One more time parameter is defined in case of each problem. It is the maximum time reduction, i.e. the maximal number of time units by which the time of each activity may be reduced. This parameter is defined either in units or in percents (in the latter case the resulting times are rounded to integer values if necessary).

---

[4] In fact, in the case of some nodes the maximum out-degree may be finally equal to $D + 1$.

[5] The floor function $\lfloor x \rfloor$ is defined as the greatest integer lower or equal to x: $\lfloor x \rfloor = \max\{y \in \mathrm{Z} \mid y \leq x\}$. E.g. $\lfloor 2.7 \rfloor = 2$, $\lfloor -2.7 \rfloor = -3$, $\lfloor 2 \rfloor = 2$, $\lfloor -2 \rfloor = -2$.

Having defined it, one may define all the possible durations for each activity. For each duration the reduction cost is generated in the following way. The starting reduction cost is a real number chosen uniformly at random from a given range. Then for each duration of the activity the increase of the cost (in comparison to the cost for the previous considered duration) is a real number chosen from another range. As both ends of the ranges have to be positive, the shortening cost sequences are always non-decreasing. In the case of the deterministic network the pair of ranges under consideration is defined separately for each type of arcs: (v_{2j+1}, v_{2j+2}), (v_{2j}, v_{2j+1}) and (v_j, v_{j+2}). In the case of the random ones, there is one range defined for all the arcs.

Now let us analyze in details the generated problems.

In the case of deterministic type network, we have tested only the problems on 10 nodes. We considered four types of test problems, distinguished with the time and basic costs generation schemes. Times on the edges $(k, k + 1)$ were equal to 5 in two first cases (Tables 2 and 3) and chosen randomly from the range [1; 5] in the remaining two (Tables 4 and 5); times on the edges $(k, k + 2)$ equal to 9 and 5 or chosen from [5; 10] and [1; 5], respectively. The basic costs for the arcs $(2k, 2k + 1)$, $(2k - 1, 2k)$ and $(k, k + 2)$ have been chosen respectively from the ranges [1; 2], [2.1; 5], [25; 50] (first and second type) and [1; 5], [1; 5], [1; 5] (third and fourth type), see Figure 3. In all four cases we considered the problems with constant and increasing costs. All the increments were chosen from some intervals as in the random case (the ranges are given in respective tables). In each case two shortening modes were tested.

In the case of the random and locally random networks we have tested the problems on 10 and 50 vertices. The problems have been tested for all the types of the ranges for activities duration (constructed using the linear, logarithmic, square root and power growth function). The maximum out-degree was set to 2.

The basic times (to be multiplied by respective growth functions) have been chosen randomly from the range [0; 5].

In the case of the problems on 50 nodes, the maximum activity compression were set to 2 units or 80%, i.e. the duration of each activity was allowed to be shortened by at most 2 units or at most down to 80% (see Tables 8, 9). Test problems on 10 nodes were examined also for the maximum shortening down to 60% (see Tables 6, 7).

The basic shortening costs (c_0) for each activity have been defined as chosen randomly from the range [1; 5]. For each possible unit of the shortening k the cost has been defined recursively as $c_k = c_{k-1} + d$, where the number d has been chosen every time uniformly at random from the range [0; 5].

For each combination of settings (four growth functions times two or three shortening modes) we tested each problem by calculating the Kaufmann and Desbazeille solution with comparison to the exact one. The computations have been performed for each possible integer shortening time T^d (dozens for each test problem). For each combination of settings 100 problems have been generated and tested.

Notice that problems calculated in this article, even though they always have discrete solutions, are not NP-hard.

8. EMPIRICAL RESULTS

The mean accuracy related to the deterministic problems is presented in Tables 2-5.

Table 2.

Network type: deterministic

Constant unit shortening cost (cost increment = 0).		Variant unit shortening cost. Cost increment $(2k, 2k + 1)$: [0; 2]. Cost increment $(2k - 1, 2k)$: [0; 5]. Cost increment $(k, k + 2)$: [0; 50].	
Short. 2 units	Short. 70%	Short. 2 units	Short. 70%
5.91329	5.16709	4.16203	3.90385

Nodes: 10. Time $(k, k + 1)$: 5. Time $(k, k + 2)$: 9. Basic cost $(2k, 2k + 1)$: [1; 2]. Basic cost $(2k - 1, 2k)$: [2.1; 5]. Basic cost $(k, k + 2)$: [25; 50].

Table 3.

Network type: deterministic

Constant unit shortening cost (cost increment = 0).		Variant unit shortening cost. Cost increment $(2k, 2k + 1)$: [0; 2]. Cost increment $(2k - 1, 2k)$: [0; 5]. Cost increment $(k, k + 2)$: [0; 50].	
Short. 2 units	Short. 70%	Short. 2 units	Short. 70%
0.00000	0.00000	0.00000	0.00000

Nodes: 10. Time $(k, k + 1)$: 5. Time $(k, k + 2)$: 5. Basic cost $(2k, 2k + 1)$: [1; 2]. Basic cost $(2k - 1, 2k)$: [2.1; 5]. Basic cost $(k, k + 2)$: [25; 50].

Table 4.

Network type: deterministic

Constant unit shortening cost (cost increment = 0).		Variant unit shortening cost. Cost increment $(2k, 2k + 1)$: [1; 2]. Cost increment $(2k - 1, 2k)$: [1; 2]. Cost increment $(k, k + 2)$: [1; 2].	
Short. 2 units	Short. 70%	Short. 2 units	Short. 70%
0.57990	0.03260	0.14376	0.05813

Nodes: 10. Time $(k, k + 1)$: [1; 5]. Time $(k, k + 2)$: [5; 10]. Basic cost $(2k, 2k + 1)$: [1; 5]. Basic cost $(2k - 1, 2k)$: [1; 5]. Basic cost $(k, k + 2)$: [1; 5].

Table 5.

Network type: deterministic

Constant unit shortening cost (cost increment = 0).		Variant unit shortening cost. Cost increment ($2k$, $2k + 1$): [1; 2]. Cost increment ($2k - 1$, $2k$): [1; 2]. Cost increment (k, $k + 2$): [1; 2].	
Short. 2 units	Short. 70%	Short. 2 units	Short. 70%
0.84420	0.00000	0.46760	0.00000

Nodes: 10. Time (k, $k + 1$): [1; 5]. Time (k, $k + 2$): [1; 5]. Basic cost ($2k$, $2k + 1$): [1; 5]. Basic cost ($2k - 1$, $2k$): [1; 5]. Basic cost (k, $k + 2$): [1; 5].

This time the accuracy is worse than in the case of randomly generated networks. It is mostly visible in the Table 2, where the numbers reach the level of about 5%. It means that the KDA is not as much useful in the cases where many critical and subcritical paths with a lot of common arcs may appear (see Figure 3). Observe that we defined the deterministic type of network in a very special way in order to make such a situation very likely. However, the poor accuracy of the KDA occurs not only because of the fact that networks are characterized by an extremely peculiar structure and by a huge number of critical and subcritical paths with a lot of common edges. The third characteristic of the networks analyzed in Table 2 is the quite unusual level of shortening costs for each type of activities (compare Table 2 with Tables 4-5). Let us recall this specific case. The unit shortening costs of activities $<2k, 2k + 1>$, $<2k - 1, 2k>$ and $<k, k + 2>$ are respectively equal to a, b and c where $b > a$ and $c > b\left(\frac{n}{2} - 1\right)$.

The mean accuracy for random and locally random problems is given in Tables 6-9.

Table 6.

Network type: locally random

Short. Mode	Linear growth	Power growth	Square root growth	Logarithmic growth
2 units	1.09354	0.54207	0.61875	0.68812
80%	0.00000	0.00000	0.00000	0.00000
60%	0.00052	0.00151	0.03298	0.03353

Nodes: 10. Out-degree: 2. Time: [1; 5]. Basic cost: [1; 5]. Cost increment: [0; 5].

Table 7.

Network type: globally random

Short. Mode	Linear growth	Power growth	Square root growth	Logarithmic growth
2 units	0.24885	0.97609	1.16946	0.45341
80%	0.00000	0.00000	0.00000	0.00000
60%	0.06152	0.00397	0.05415	0.00714

Nodes: 10. Out-degree: 2. Time: [1; 5]. Basic cost: [1; 5]. Cost increment: [0; 5].

Table 8.

Network type: locally random

Short. Mode	Linear growth	Power growth	Square root growth	Logarithmic growth
2 units	0.28675	0.57902	0.34920	0.31711
80%	0.03424	0.00126	0.00510	0.02037

Nodes: 50. Out-degree: 2. Time: [1; 5]. Basic cost: [1; 5]. Cost increment: [0; 5].

Table 9.

Network type: globally random

Short. Mode	Linear growth	Power growth	Square root growth	Logarithmic growth
2 units	0.02317	0.00000	0.55802	0.19056
80%	0.00000	0.00000	0.00000	0.00000

Nodes: 50. Out-degree: 2. Time: [1; 5]. Basic cost: [1; 5]. Cost increment: [0; 5].

As one can see, the accuracy is usually better in networks consisting of more nodes (compare Table 8 with Table 6 and Table 9 with Table 7). The differences between the KD solutions and the exact ones are not very significant – only few of them exceeds the level of 1%, while many of them are less than 10^{-5}%. This proves our supposition that the modified version of KDA behaves quite good in the situation where the network does not contain too much critical paths (which is very likely in our case, as the network is quite sparse).

We tested two kinds of random networks. The ones called *globally random* have arcs distributed in a sense uniformly, while the *locally random networks* have rather a structure close to the D^{th} power of path (i.e. the graph where the vertex v_i is adjacent with the vertices v_j, for which the inequality $|i - j| \leq D$ holds). We hoped that this difference in the structure of the network would lead to some substantial differences in the accuracy of the examined algorithm. The influence of the network structure on the algorithm is not obvious.

In the case of logarithmic growth, the algorithm performs better for globally random networks. In the case of square root growth, it acts better when networks are locally random (with one exception, however in this case – 50 nodes, 80% shortening – the difference is hardly visible). In the case of power growth the algorithm performs better on small locally random and bigger globally random networks. In the case of linear growth, again the algorithm is better for globally random networks (the only exception seems to be not significant). As we can see, in most cases the more ordered structure of the network (locally random) implies a worse performance of the KDA. However, this does not mean that the structure itself has the influence on the way the KDA works. It has to be analyzed together with other features of the problem, in particular the distribution of the shortening costs.

9. CONCLUSIONS

The goal of our article was to check the accuracy of the Kaufmann and Desbazeille algorithm for the worst case described in Anholcer, Gaspars-Wieloch (2011) and for randomly generated problems. As one can notice the difference between the exactitude of the procedure for the deterministic and random networks is quite significant. The KDA is essentially inefficient only in specific cases. Nevertheless we should remember that the computational experiments focused on the modified version of the KDA. The accuracy of the original method is certainly worse.

Poznan University of Economics

REFERENCES

[1] Anholcer M., Gaspars-Wieloch H., (2011), The Efficiency Analysis of the Kaufmann and Desbazeille Algorithm for the Deadline Problem, *Operations Research and Decisions* 2011/2, Wrocław.
[2] Bladowski S., (1970), *Metody sieciowe w planowaniu i organizacji pracy*, PWE, Warszawa.
[3] Bozarth C., Handfield R.B., (2005), *Introduction to Operations and Supply Chain Management*, Prentice Hall, Pearson.
[4] Cormen T. H., Leiserson C. E., Rivest R. L., (2001), *Introduction to Algorithms*, MIT Press, Cambridge.
[5] De P., Dunne E. J., Gosh J. B., Wells C. E., (1995), The Discrete Time-Cost Tradeoff Problem Revisited, *European Journal of Operations Research*, 81, 225–238.
[6] De P., Dunne E. J., Gosh J. B., Wells C. E., (1997), Complexity of the Discrete Time-Cost Trade-Off Problem for Project Networks, *Operations Research*, 45, 302–306.
[7] Edmonds J., Karp R. M., (1972), Theoretical Improvements in the Algorithmic Efficiency for Network Flow Problems, *Journal of the ACM*, 19, 248–264.
[8] Ford L. R., Fulkerson D. R, (1962), *Flows in Networks*, Princeton University Press, Princeton, N. J.
[9] Fusek A., Nowak K., Podleski H., (1967), *Analiza drogi krytycznej* (*CPM i PERT*). *Instrukcja programowana*, PWE, Warszawa.

[10] Gaspars H., (2006), Analiza czasowo-kosztowa (CPM-COST). Algorytm a model optymalizacyjny, *Badania Operacyjne i Decyzje*, 2006, nr 1, Wydawnictwo Politechniki Wrocławskiej, 5–19.
[11] Gaspars H., (2006), Propozycja nowego algorytmu w analizie czasowo-kosztowej przedsięwzięć, *Badania Operacyjne i Decyzje*, 2006, nr 3–4, Wydawnictwo Politechniki Wrocławskiej, 5–27.
[12] Gaspars-Wieloch H., (2009), *Metody optymalizacji czasowo-kosztowej przedsięwzięcia*, (doctoral thesis), Uniwersytet Ekonomiczny w Poznaniu, Poznań.
[13] Gaspars-Wieloch H., (2008), Przegląd wybranych metod skracania czasu realizacji przedsięwzięcia, w: Kopańska-Bródka *D*., (red.)*, Metody i zastosowania badań operacyjnych*, Wydawnictwo Akademii Ekonomicznej w Katowicach.
[14] Gedymin O., (1974), *Metody optymalizacji w planowaniu sieciowym*, PWN, Warszawa.
[15] Giard V., (1991), *Gestion de Projets*, Economica, Paris.
[16] Hendrickson C., Au T., (1989), *Project Management for Construction. Fundamental Concepts for Owners, Engineers, Architects and Builders*, Prentice Hall.
[17] Idźkiewicz A. Z., (1967), *PERT. Metody analizy sieciowej*, PWN, Warszawa.
[18] Kaufmann A., Desbazeille G., Ventura E., (1964), *La Méthode du Chemin Critique*, Dunod, Paris.
[19] Kopańska-Bródka D., (1998), *Wprowadzenie do badań operacyjnych,* wydanie drugie poprawione, Wydawnictwo Akademii Ekonomicznej w Katowicach.
[20] Korzan B., (1978), *Elementy teorii grafów i sieci. Metody i zastosowania*, Wydawnictwa Naukowo-Techniczne, Warszawa.
[21] Kukuła K. (red.), Jędrzejczyk Z., Skrzypek J., Walkosz A., (1996), *Badania operacyjne w przykładach i zadaniach*, PWN, Warszawa.
[22] Muller Y., (1965), *Exercices d'Organisation et de la Recherche Opérationnelle*, Eyrolles, Paris.
[23] Muller Y., (1964), *Organisation et Recherche Opérationnelle*, Paris, Eyrolles.
[24] Panagiotakopoulos D., (1977), A CPM Time-Cost Computational Algorithm for Arbitrary Activity Cost Functions, *INFOR*, 15 (2), 183–195.
[25] Schrijver A., (2003), *Combinatorial Optimization: Polyhedra and Efficiency*, Springer – Verlag Berlin Heidelberg.
[26] Siudak M., (1998), *Badania operacyjne*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
[27] Skutella M., (1998), Approximation Algorithms for the Discrete Time-Cost Tradeoff Problem*, Mathematics of Operations Research*, 23 (4), 909–928.
[28] Trocki M. (red.)*,* Grucza B., Ogonek K., (2003), *Zarządzanie projektami*, PWE, Warszawa.
[29] Waters D., (1998), *A Practical Introduction to Management Science*, Second edition, Addison-Wesley Longman.

DOKŁADNOŚĆ ALGORYTMU KAUFMANNA I DESBAZEILLE W PROBLEMACH OPTYMALIZACJI CZASOWO-KOSZTOWEJ PROJEKTU

Streszczenie

Analiza czasowo–kosztowa jest bardzo ważnym elementem zarządzania projektem. Algorytm Kaufmanna–Desbazeille dla tego problemu jest przez wielu autorów określany mianem dokładnego, lecz w kilku pracach wykazano, iż w niektórych przypadkach stosowanie tej procedury prowadzi jedynie do rozwiązań bliskich optimum. W artykule wyznaczamy średnią dokładność algorytmu dla pewnej liczby sieci o z góry ustalonej bądź losowo wygenerowanej strukturze. Dokładność procedury Kaufmanna i Desbazeille jest najniższa, gdy:

– sieć jest generowana w sposób deterministyczny (parzysta liczba węzłów, sieć składa się z samych łuków łączących sąsiednie węzły, sąsiednie węzły parzyste i sąsiednie węzły nieparzyste, a więc posiada wiele ścieżek krytycznych i podkrytycznych ze wspólnymi łukami),
– każdy typ czynności w tak skonstruowanej sieci ma bardzo specyficzne charakterystyki czasowo-kosztowe.

Struktura sieci ma wpływ na wydajność algorytmu. Powinna być jednak analizowana łącznie z rozkładem jednostkowych kosztów skrócenia czynności.

Słowa kluczowe: analiza czasowo-kosztowa projektów, sieć, ścieżka krytyczna, dokładność algorytmu, skracanie czasu realizacji projektu, krzywe czasowo-kosztowe, minimalizacja kosztu przy zadanym czasie dyrektywnym

ACCURACY OF THE KAUFMANN AND DESBAZEILLE ALGORITHM FOR TIME-COST TRADE-OFF PROJECT PROBLEMS

Abstract

The time-cost tradeoff analysis is a very important issue in the project management. The Kaufmann-Desbazeille method is considered by numerous authors as an exact algorithm to solve that problem, but in some articles it has been proved that for specific network cases the procedure only leads to quasi-optimal solutions. In this paper we calculate the average accuracy of the algorithm for several deterministic and randomly generated networks. The accuracy of the KDA is the worst when:

– the network is generated in a deterministic way (an even number of nodes, the network contains only arcs connecting neighbouring nodes, neighbouring even nodes and neighbouring odd nodes, thus it has many critical and subcritical paths with a lot of common arcs),
– each type of activities in such a network has very specific time-cost characteristics.

The structure of the network has the influence on the performance of KDA. It should be however analyzed together with the distribution of the shortening costs.

Keywords: time-cost tradeoff project analysis (TCTP-analysis), network, critical path, accuracy of the algorithm, project compression time, time-cost curves, deadline problem

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

RENATA WRÓBEL-ROTTER

ESTYMOWANE MODELE RÓWNOWAGI OGÓLNEJ I AUTOREGRESJA WEKTOROWA. ASPEKTY TEORETYCZNE[1]

1. WSTĘP

Model ekonometryczny rozważany w pracy powstaje poprzez przyjęcie rozkładu *a priori,* generowanego przez estymowany model równowagi ogólnej, dla wektorowej autoregresji bez restrykcji, prowadząc do klasy hybrydowych wektorowych autoregresji znanych w literaturze pod pojęciem DSGE-VAR (ang. *Dynamic Stochastic General Equilibrium Vector AutoRegression*). Łączną one zalety stylizowanych strukturalnych modeli makroekonomicznych z elastycznością wektorowej autoregresji, pozwalając danym empirycznym wybrać stopień w jakim potwierdzają one każde z podejść. Mogą one być wykorzystywane do oceny poprawności specyfikacji konstrukcji wywodzącej się z teorii ekonomii, dopuszczając jednocześnie możliwość uchylenia restrykcji z niej wynikających, które nie pozostają w zgodzie z danymi obserwowanymi. Celem pracy jest kompleksowe omówienie i podsumowanie zagadnień teoretycznych związanym z modelami DSGE-VAR, m.in. szczegółowa prezentacja rozkładów prawdopodobieństwa leżących u podstaw konstrukcji modelu połączonego, omówienie roli parametru wagowego i detali związanych z konstrukcją rozkładu a priori, w tym wyprowadzenie kompletnych formuł na momenty teoretyczne zmiennych stanu, oraz prezentacja kolejnych etapów budowy modelu. Praca ma charakter teoretyczny i jest autorską syntezą informacji związanych z metodologią DSGE-VAR. Wypełnia ona lukę w literaturze polskiej przedstawiając całościowe ujęcie modelu i może stanowić użyteczne wprowadzenie do badań empirycznych.

[1] Praca wykonana w ramach badań statutowych Katedry Ekonometrii i Badań Operacyjnych Uniwersytetu Ekonomicznego w Krakowie. Autorka pragnie złożyć podziękowania Profesorowi Jackowi Osiewalskiemu oraz uczestnikom seminarium Katedry Ekonometrii i Badań Operacyjnych za komentarze i dyskusję podczas prezentacji opracowania.

2. OGÓLNA CHARAKTERYSTYKA

Metodologia pozwalająca na połączenie wnioskowania na podstawie estymowanych modeli równowagi ogólnej z modelami wektorowej autoregresji została zaproponowana w pracy Del Negro i Schorfheide (2004) i następnie rozwinięta przez Del Negro i in. (2007). Określa ona klasę hybrydowych modeli wektorowej autoregresji (DSGE-VAR), które powstały w wyniku poszukiwania metod uwzględniania w modelach wektorowej autoregresji informacji wstępnych, mających za zadanie ich powiązanie z teorią ekonomii i poprawienie ich własności. Pierwotnie prace te koncentrowały się na zastosowaniu modeli strukturalnych, podbudowanych teorią ekonomii, do konstrukcji rozkładów *a priori* dla wektorowej autoregresji (por. np. Ingram i Whiteman, 1994), bądź do porównań własności statystycznych makroekonomicznych szeregów czasowych (por. np. DeJong i in., 1996). Modele teoretyczne dostarczały porównywalnego rozkładu *a priori*, ocenianego w kategoriach zdolności prognostycznej uzyskanej postaci hybrydowej, jak metodologia, w której zmienne wektorowej autoregresji są traktowane *a priori* jako grupa niezależnych procesów błądzenia losowego, zaproponowana w pracy Doana i in. (1983) i Littermana (1986). Pierwszy tego typu model, estymowany metodą największej wiarygodności, powstał po zdefiniowaniu procesu autoregresyjnego dla wektora zakłóceń losowych w równaniu obserwacji, reprezentacji modelu równowagi ogólnej w przestrzeni stanów, i został zaproponowany w pracy Irelanda (2004). Dalszy rozwój metodologii polegał na konstrukcji rozkładów prawdopodobieństwa, które w sposób formalny łączyły wnioskowanie na podstawie wektorowej autoregresji z modelami posiadającymi uzasadnienie w teorii ekonomii. Techniki te pozwoliły na opracowane metod formalnego wnioskowania o parametrach modelu równowagi ogólnej na podstawie wektorowej autoregresji w modelu połączonym (por. np. Del Negro i Schorfheide, 2004; Del Negro i in., 2007). Z obecnie opracowanych zastosowań można wymienić prace m.in. Lee i in. (2007), Liu i Gupty (2008), Watanabe (2007), Chowa i McNelisa (2010), Adolfsona i in. (2008), Kolasy i in. (2012) i Brzoza-Brzeziny i Kolasy (2012). Niniejsza praca jest poświęcona szczegółowemu i kompleksowemu omówieniu metodologii DSGE-VAR, która jest niezbędna do prezentacji zagadnień empirycznych, zawartych w opracowaniu: Wróbel-Rotter (2013b). Obydwie prace stanowią kontynuację badań związanych ze stosowaniem estymowanych modeli równowagi ogólnej (DSGE) w praktyce, w ramach których szczegółowo omówiono podstawy teoretyczne wraz z wyprowadzeniem równań strukturalnych: Wróbel-Rotter (2011a), (2011c), (2012b), sposoby doboru rozkładu a priori dla wag i parametru wagowego wraz z omówieniem ich wpływu na funkcje odpowiedzi impulsowych: Wróbel-Rotter (2013c), (2013d) oraz aspekty estymacyjne i numeryczne: Wróbel-Rotter (2011b), (2012a), (2013a). Badanie te poprzedza ogólna charakterystyka estymowanych modeli równowagi ogólnej, zawarta w pracach Wróbel-Rotter (2012c), (2012d) oraz we wcześniejszych: Wróbel-Rotter (2007c), (2007a), (2007b), (2008).

Model DSGE-VAR składa się z pomocniczego modelu wektorowej autoregresji, służącego aproksymacji rozwiązania zlinearyzowanego modelu równowagi ogólnej i konstrukcji rozkładu *a priori*, oraz zasadniczego modelu wektorowej autoregresji, szacowanego na danych rzeczywistych. Może on być interpretowany jako identyfikowalny model wektorowej autoregresji (ang. *identified* VAR), nie zaś jako forma zredukowana modelu strukturalnego (por. np. An i Schorfheide, 2007). Modele DSGE-VAR dostarczają narzędzia służącego odpowiedzi na pytanie w jakim stopniu dane empiryczne potwierdzają hipotezy wnoszone przez model równowagi ogólnej a w jakim są one lepiej opisywane przez model wektorowej autoregresji bez ograniczeń. Stosowanie modelu strukturalnego jako punktu odniesienia dla procesów wektorowej autoregresji zakłada, że teoria ekonomii, ujęta poprzez stylizowane zależności zdefiniowane w teoretycznej gospodarce, nadaje się do jej uwzględnienia do opisu danych rzeczywistych. Model DSGE-VAR można postrzegać jako sposób na poprawienie własności wektorowej autoregresji, poprzez uwzględnienie informacji wstępnej, wynikającej z teorii ekonomii, bądź też jako technikę umożliwiającą złagodzenie restrykcji obecnych w modelu równowagi ogólnej i ocenę poprawności jego specyfikacji. Budowa modelu DSGE-VAR jest procesem hierarchicznym, zaczynającym się od specyfikacji rozkładu *a priori* dla wektora parametrów modelu równowagi ogólnej, warunkowo względem którego definiuje się rozkład *a priori* dla współczynników wektorowej autoregresji. Umożliwia to, po uwzględnieniu funkcji wiarygodności, wnioskowanie *a posteriori* zarówno o parametrach modelu strukturalnego jak i współczynnikach wektorowej autoregresji. Koncepcja budowy rozkładu *a priori* bazuje na idei dolosowania danych, które następnie służą estymacji modelu pomocniczego pozwalającego na przekazanie informacji wstępnej; podejście takie stosowali m.in. Sims i Zha (1998). Ma ono również swoje uzasadnienie w pionierskich pracach z zakresu łączenia w modelu statystycznym wiedzy spoza próby i informacji niesionej przez obserwacje oraz metodach ich estymacji, zaproponowane przez Theila i Goldbergera (1961). Koncepcja modeli pomocniczych jest związana również z bayesowskimi metodami wnioskowania nie wprost (ang. *indirect inference*), (por. np. Gallant i McCulloch, 2009).

3. MODEL POŁĄCZONY

Łączne wnioskowanie o współczynnikach i macierzy kowariancji wektorowej autoregresji: Φ i Σ_u oraz o parametrach θ modelu równowagi ogólnej jest możliwe po zdefiniowaniu modelu połączonego, który jest określony przez model wektorowej autoregresji aproksymujący rozwiązanie zlinearyzowanej postaci modelu strukturalnego oraz wektorową autoregresję bez restrykcji dla danych obserwowalnych. Powstały w ten sposób model połączony wektorowej autoregresji umożliwia, na gruncie metod bayesowskich, jednoczesne wnioskowanie o Φ, Σ_u i θ, przy ustalonym wektorze obserwacji. Łączny rozkład *a priori* dla Φ, Σ_u i θ jest budowany w sposób

hierarchiczny, poprzez założenie rozkładu *a priori* dla parametrów modelu równowagi ogólnej, a następnie, warunkowo względem niego, przyjęcie rozkładu *a priori* dla współczynników wektorowej autoregresji:

$$p(\Phi,\Sigma_u,\theta) = p(\Phi,\Sigma_u|\theta)p(\theta), \tag{1}$$

gdzie $p(\Phi,\Sigma_u|\theta)$ to rozkład *a priori* współczynników i macierzy kowariancji wektorowej autoregresji bez restrykcji, warunkowy względem θ, $p(\theta)$ jest brzegowym rozkładem *a priori* dla parametrów modelu równowagi ogólnej, specyfikowanym w sposób standardowy, na podstawie oczekiwań co do jego własności i zakresu możliwych wartości parametrów w danym przypadku empirycznym.

Statystyczny model bayesowski, zdefiniowany poprzez łączny rozkład wektora obserwacji Y, parametrów wektorowej autoregresji i modelu strukturalnego ma postać:

$$p(Y,\Phi,\Sigma_u,\theta) = p(Y|\Phi,\Sigma_u,\theta)p(\Phi,\Sigma_u|\theta)p(\theta), \tag{2}$$

gdzie warunkowa gęstość obserwacji: $p(Y|\Phi,\Sigma_u,\theta) = p(Y|\Phi,\Sigma_u)$ i funkcja wiarygodności: $|(\Phi, \Sigma_u,| Y)$ nie zależą od θ. Łączny rozkład *a posteriori* w modelu hybrydowym współczynników wektorowej autoregresji i parametrów modelu równowagi ogólnej otrzymujemy z wzoru Bayesa:

$$p(\Phi,\Sigma_u,\theta|Y) = \frac{p(Y,\Phi,\Sigma_u,\theta)}{p(Y)}, \tag{3}$$

,

gdzie $p(Y)$ oznacza brzegową gęstość obserwacji, przy czym zachodzą równości:

$$\frac{p(Y,\Phi,\Sigma_u,\theta)}{p(Y)} = \frac{p(Y|\Phi,\Sigma_u)p(\Phi,\Sigma_u,\theta)}{p(Y)} = \frac{p(Y|\Phi,\Sigma_u)p(\Phi,\Sigma_u|\theta)p(\theta)}{p(Y)}. \tag{4}$$

Wnioskowania *a posteriori* z modelu hybrydowego, o parametrach wektorowej autoregresji, warunkowo względem parametrów θ modelu strukturalnego, dokonuje się po faktoryzacji łącznego rozkładu *a posteriori*:

$$p(\Phi,\Sigma_u,\theta|Y) = p(\Phi,\Sigma_u|Y,\theta)p(\theta|Y), \tag{5}$$

gdzie $p(\Phi,\Sigma_u|Y,\theta)$ jest rozkładem *a posteriori* współczynników i macierzy kowariancji wektorowej autoregresji, warunkowym względem parametrów modelu równowagi

ogólnej, $p(Y|\theta)$ jest brzegowym względem Φ i Σ_u rozkładem *a posteriori* parametrów modelu równowagi ogólnej. Charakterystyki warunkowego rozkładu *a posteriori* $p(\Phi,\Sigma_u|Y,\theta)$ stanowią podstawę do analiz ekonomicznych, uzyskanych z hybrydowego modelu wektorowej autoregresji.

Łączny rozkład *a posteriori* $p(\Phi,\Sigma_u,\theta|Y)$ może zostać poddany dekompozycji umożliwiającej wnioskowanie o parametrach modelu strukturalnego. Warunkowy rozkład $p(\theta|\Phi, \Sigma_u,Y)$ *a posteriori* parametrów modelu równowagi ogólnej, przy ustalonych Φ i Σ_u uzyskuje się z modelu hybrydowego po zapisaniu:

$$p(\Phi,\Sigma_u,\theta|Y) = p(\theta|\Phi,\Sigma_u,Y)\,p(\Phi,\Sigma_u|Y), \tag{6}$$

gdzie $p(\Phi,\Sigma_u|Y)$ jest rozkładem *a posteriori* współczynników wektorowej autoregresji, brzegowym względem θ, natomiast warunkowy względem Φ i Σ_u rozkład *a posteriori* dla θ nie zależy od wektora obserwacji: $p(\theta|\Phi, \Sigma_u,Y)= p(\theta|\Phi, \Sigma_u)$. Oznacza to, że wnioskowanie, z modelu hybrydowego, o parametrach modelu strukturalnego na podstawie danych rzeczywistych jest możliwe pośrednio, poprzez wnioskowanie o parametrach Φ oraz Σ_u wektorowej autoregresji.

Brzegowy rozkład *a posteriori* $p(\theta|Y)$, wektora parametrów modelu strukturalnego θ, zawierający dostępne po zaobserwowaniu danych Y informacje, jest uzyskiwany na podstawie wzoru Bayesa: $p(\theta|Y) = p(Y|\theta)p(\theta)/p(Y)$, gdzie $p(Y) = \int p(Y|\theta)p(\theta)/d\theta$ jest brzegową, względem wszystkich parametrów modelu hybrydowego, gęstością obserwacji. Gęstość obserwacji w modelu połączonym, warunkowa względem θ i brzegowa względem parametrów Φ i Σ_u, jest dana przez:

$$\begin{aligned} p(Y|\theta) &= \frac{p(Y,\theta)}{p(\theta)} = \frac{\int p(Y|\Phi,\Sigma_u,\theta)\,p(\Phi,\Sigma_u|\theta)\,p(\theta)d(\Phi,\Sigma_u)}{p(\theta)} = \\ &= \int p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta)d(\Phi,\Sigma_u) = \frac{p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta)}{p(\Phi,\Sigma_u|Y)}. \end{aligned} \tag{7}$$

Funkcja gęstości łącznego rozkładu *a posteriori* $p(\theta|Y)$ parametrów modelu strukturalnego jest proporcjonalna do iloczynu gęstości *a priori p(θ)* i warunkowej gęstości obserwacji $p(Y|\theta)$, traktowanej jako funkcja θ, przy ustalonym Y, i oznaczanej przez $l(\theta|Y)$: $p(\theta|Y) \propto l(\theta|Y)\,p(\theta)$. Alternatywnie, brzegowy rozkład *a posteriori* $p(\theta|Y)$, warunkowy względem wektora obserwacji Y, dla parametrów modelu równowagi ogólnej, uzyskuje się z łącznego rozkładu *a posteriori*: $p(\Phi,\Sigma_u,\theta|Y)$, współczynników wektorowej autoregresji Φ i Σ_u oraz parametrów θ modelu strukturalnego, poprzez obliczenie całki wielowymiarowej:

$$p(\theta|Y) = \int p(\Phi,\Sigma_u,\theta|Y)d(\Phi,\Sigma_u) = \int p(\theta|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|Y)d(\Phi,\Sigma_u). \tag{8}$$

4. MODEL STRUKTURALNY I WEKTOROWA AUTOREGRESJA

Model wektorowej autoregresji rzędu p bez restrykcji (VAR(p)), zapisany dla n obserwowanych zmiennych w momencie t, ma postać:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ...\Phi_p Y_{t-p} + u_t, \qquad u_t \sim N^{(n)}(0, \Sigma_u), \tag{9}$$

gdzie Y_t oznacza kolumnę $(n \times 1)$ zmiennych obserwowalnych, Φ_i dla $i=1,...,p$, są macierzami $(n \times n)$ zawierającymi współczynniki wektorowej autoregresji, Φ_0 jest wektorem $(n \times 1)$ wyrazów wolnych, u_t jest wektorem $(n \times 1)$ zakłóceń losowych postaci zredukowanej, $N^{(n)}(0, \Sigma_u)$ oznacza n-wymiarowy rozkład normalny, o wektorze wartości oczekiwanych równym wektorowi zerowemu, $E(u_t) = 0$, i macierzy kowariancji $E(u_t u_t') = \Sigma_u$, o wymiarach $(n \times n)$, U_t i X_t są niezależne, $E(u_t u_{t-j}') = 0$. W zapisie macierzowym otrzymujemy: $Y = X\Phi + U$; gdzie Y jest macierzą $(T \times n)$ składającą się z wierszy Y_t', X oznacza macierz $(T \times (1+np))$ o wierszach $X_t' = [1\ Y_{t-1}'\ Y_{t-2}' \ldots Y_{t-p}']$, U jest macierzą $(T \times n)$ składającą się z wierszy u_t', $\Phi = \Phi'_0\ \Phi'_1 \ldots \Phi'_p]'$ jest macierzą $((1+np) \times n)$ współczynników wektorowej autoregresji. Założenie n-wymiarowych rozkładów normalnych dla u_t, oznacza, że przy ustalonym wektorze wartości początkowych $Y_{1-p}, ..., Y_0$, warunkowa gęstość obserwacji ma postać:

$$p(Y|\Phi, \Sigma_u) = (2\pi)^{-nT/2} \det(\Sigma_u)^{-T/2} \exp\{-0.5tr[\Sigma_u^{-1}(Y - X\Phi)'(Y - X\Phi)]\}, \tag{10}$$

gdzie $tr(.)$ to ślad macierzy. Łączny rozkład *a posteriori* otrzymujemy z wzoru Bayesa:

$$p(\Phi, \Sigma_u|Y) = \frac{p(Y|\Phi, \Sigma_u)\, p(\Phi, \Sigma_u)}{\int p(Y|\Phi, \Sigma_u)\, p(\Phi, \Sigma_u)\, d(\Phi, \Sigma_u)} \tag{11}$$

gdzie w mianowniku mamy brzegową gęstość obserwacji. Funkcja gęstości rozkładu *a posteriori* $p(\Phi,\Sigma_u|Y)$ jest proporcjonalna do iloczynu funkcji wiarygodności $|(\Phi,\Sigma_u|Y)$, będącej warunkową gęstością obserwacji $p(Y|\Phi,\Sigma_u)$, traktowaną jako funkcja parametrów Φ i Σ_u, oraz funkcji gęstości rozkładu *a priori* $p(\Phi,\Sigma_u)$ dla macierzy współczynników i macierzy kowariancji:

$$p(\Phi, \Sigma_u|Y) \propto \ell(\Phi, \Sigma_u|Y)\, p(\Phi, \Sigma_u). \tag{12}$$

Specyfikacja rozkładu *a priori* $p(\Phi,\Sigma_u) = p(\Phi,\Sigma_u|\theta)$ odgrywająca kluczową rolę we wnioskowaniu na podstawie modelu hybrydowego wektorowej autoregresji, jest

oparta na estymowanym modelu równowagi ogólnej (modelu strukturalnym), którego równania są zapisywane w formie reprezentacji w przestrzeni stanów:

$$s_t = As_{t-1} + B\varepsilon_t \text{ oraz } Y_t = F + Cs_t + v_t, \tag{13}$$

gdzie s_t oznacza wektor stanu, elementy macierzy A i B są nieliniowymi funkcjami θ, wynikającymi z rozwiązania postaci zlinearyzowanej modelu strukturalnego, v_t jest wektorem zakłóceń losowych w równaniu obserwacji, najczęściej o rozkładzie normalnym, $v_t \sim N^{(n)}(0, \Sigma_v)$, ε_t oznacza wektor innowacji związanych z egzogenicznymi zakłóceniami losowymi występującymi w postaci strukturalnej, $\varepsilon_t \sim N^{(n^*)}(0, \Sigma_\varepsilon)$, F i C oznaczają wektor i macierz znanych stałych bądź parametrów podlegających oszacowaniu, Y_t jest procesem obserwowalnym i kowariancyjnie stacjonarnym, co wynika z konstrukcji modelu.

Model równowagi ogólnej jest tutaj traktowany jako podstawa do konstrukcji rozkładu *a priori*, poprzez który są uwzględniane w wektorowej autoregresji restrykcje, wynikające z przyjętych do jego konstrukcji założeń. Budowa rozkładu *a priori* wykorzystuje formalne wnioskowanie na podstawie pomocniczego modelu wektorowej autoregresji, przybliżającego model strukturalny. Aproksymacja rozwiązania liniowej postaci modelu równowagi ogólnej wektorową autoregresją jest związana z wyznaczeniem szeregu restrykcji, wiążących jej współczynniki, wynikających z postaci modelu strukturalnego. Możliwym sposobem transformacji rozwiązania zlinearyzowanej postaci modelu strukturalnego w formę modelu wektorowej autoregresji jest rozpatrzenie odwzorowania wektora parametrów θ w macierz współczynników $\tilde{\Phi}$ i macierz kowariancji $\tilde{\Sigma}_u$ dla pomocniczej autoregresji:

$$\tilde{Y} = \tilde{X}\tilde{\Phi} + \tilde{U}, \tag{14}$$

która dla jednej obserwacji wektorowej ma postać:

$$\tilde{Y}_t = \tilde{\Phi}_0 + \sum_{i=1}^{p} \tilde{\Phi}_i \tilde{Y}_{t-i} + \tilde{u}_t = \tilde{\Phi}' \tilde{X}_t + \tilde{u}_t, \quad t = 1,...,\tilde{T}, \quad \tilde{u}_t \sim iidN^{(n)}(0, \tilde{\Sigma}_u), \tag{15}$$

gdzie $\tilde{Y}_t$ jest wektorem *(n×1)* zawierającym zmienne obserwowalne, które odpowiadają zmiennym endogenicznym modelu równowagi ogólnej przyjętym do jego estymacji w równaniu obserwacji, macierz $\tilde{Y}$, o wymiarach *($\tilde{T}$ ×n)*, składa się z wierszy $\tilde{Y}_t'$, $\tilde{T}$ oznacza liczebność teoretycznej próby, $\tilde{\Phi}' = [\tilde{\Phi}_0' \; \tilde{\Phi}_1' \; ... \; \tilde{\Phi}_p']$, $\tilde{\Phi}$ jest macierzą *((1+np)×n)* współczynników autoregresji, $\tilde{X}$ oznacza macierz *($\tilde{T}$ ×(1+np))* o wierszach $\tilde{X}_t' = [1 \; \tilde{Y}_{t-1}' \; \tilde{Y}_{t-2}' ... \tilde{Y}_{t-p}']$, $\tilde{X}_t$, $\tilde{X}_t$ jest wektorem *((1+np)×1)*, $\tilde{U}$ jest macierzą *($\tilde{T}$ ×ũn)* składającą się z wierszy $\tilde{u}_t'$, zaś $\tilde{u}_t$ jest wektorem *(n×1)* zakłóceń losowych

o rozkładzie normalnym, z zerowym wektorem wartości oczekiwanych i macierzą kowariancji $\widetilde{\Sigma}_u$. Funkcja wiarygodności w tym przypadku ma standardową postać:

$$\begin{aligned}\ell(\widetilde{\Phi},\widetilde{\Sigma}_u\left|\widetilde{Y}\right.) &= (2\pi)^{-n\widetilde{T}/2}\det(\widetilde{\Sigma}_u)^{-\widetilde{T}/2}\exp\{-0.5tr[\widetilde{\Sigma}_u^{-1}(\widetilde{Y}-\widetilde{X}\widetilde{\Phi})'(\widetilde{Y}-\widetilde{X}\widetilde{\Phi})]\} = \\ &= (2\pi)^{-n\widetilde{T}/2}\det(\widetilde{\Sigma}_u)^{-\widetilde{T}/2}\exp\{-0.5tr[\widetilde{\Sigma}_u^{-1}(\widetilde{Y}'\widetilde{Y}-(\widetilde{X}\widetilde{\Phi})'\widetilde{Y}-\widetilde{Y}'\widetilde{X}\widetilde{\Phi}+(\widetilde{X}\widetilde{\Phi})'\widetilde{X}\widetilde{\Phi}]\}.\end{aligned} \tag{16}$$

Odwzorowanie parametrów strukturalnych θ w macierze $\widetilde{\Phi}$ i $\widetilde{\Sigma}_u$ uzyskuje się poprzez formalne wnioskowanie (ang. *population regression*) z wykorzystaniem teoretycznych momentów pierwszego i drugiego rzędu zmiennych $\widetilde{Y}_t$ i $\widetilde{X}_t$. Są one bezpośrednio zależne od parametrów modelu równowagi ogólnej i wyrażają się wartościami oczekiwanymi $E_\theta(.)$, wyznaczanymi względem rozkładu prawdopodobieństwa θ: $\Gamma_{xx}^*(\theta)=E_\theta(\widetilde{X}_t\widetilde{X}_t')$, $\Gamma_{xy}^*(\theta)=E_\theta(\widetilde{X}_t\widetilde{Y}_t')$, $\Gamma_{yy}^*(\theta)=E_\theta(\widetilde{Y}_t\widetilde{Y}_t')$ i $\Gamma_{yx}^*(\theta)=\Gamma_{xy}^*{}'(\theta)$. Funkcja wiarygodności zapisana dla przeskalowanych momentów teoretycznych, zamiast momentów empirycznych $\widetilde{Y}_t$ i $\widetilde{X}_t$:

$$\ell^*(\Phi,\Sigma_u\left|\widetilde{Y}\right.)\propto\det(\Sigma_u)^{-n\widetilde{T}/2}\exp\{-0.5tr[\widetilde{T}\widetilde{\Sigma}_u^{-1}(\Gamma_{yy}^*(\theta)-\widetilde{\Phi}'\Gamma_{xy}^*(\theta)-\Gamma_{yx}^*(\theta)\widetilde{\Phi}+\widetilde{\Phi}'\Gamma_{xx}^*(\theta)\widetilde{\Phi}]\}, \tag{17}$$

prowadzi, przy założeniu *a priori* nieinformacyjnego rozkładu Jeffreysa dla macierzy współczynników $\widetilde{\Phi}$ i macierzy kowariancji $\widetilde{\Sigma}_u$: $p(\widetilde{\Phi},\widetilde{\Sigma}_u)\propto\det(\widetilde{\Sigma}_u)^{-(n+1)/2}$ do gęstości łącznego rozkładu *a posteriori*:

$$p(\widetilde{\Phi},\widetilde{\Sigma}_u\left|\theta\right.)\propto\det(\widetilde{\Sigma}_u)^{-(\widetilde{T}+n+1)/2}\exp\{-0.5tr[\widetilde{T}\,\widetilde{\Sigma}_u^{-1}(\Gamma_{yy}^*(\theta)-\widetilde{\Phi}'\Gamma_{xy}^*(\theta)-\Gamma_{yx}^*(\theta)\widetilde{\Phi}+\widetilde{\Phi}'\Gamma_{xx}^*(\theta)\widetilde{\Phi})]\}, \tag{18}$$

która jest przyjmowana w miejsce gęstosci *a priori* współczynników hybrydowej wektorowej autoregresji: $p(\Phi,\Sigma_u\left|\theta\right.)=p(\widetilde{\Phi},\widetilde{\Sigma}_u\left|\theta\right.)$. Brzegowy rozkład *a posteriori* macierzy kowariancji $\widetilde{\Sigma}_u$, przy danym θ, posiada formę odwróconego rozkładu Wisharta, (por. np. Zellner, 1971):

$$p(\widetilde{\Sigma}_u\left|\theta\right.)=f_{IW}(\widetilde{\Sigma}_u\left|\widetilde{T}\hat{\widetilde{\Sigma}}_u(\theta),\widetilde{T}-(1+np),n\right.), \tag{19}$$

natomiast warunkowy względem $\widetilde{\Sigma}_u$ i θ rozkład *a posteriori* dla współczynników autoregresji $\widetilde{\Phi}$, ma postać rozkładu macierzowego normalnego:

$$p(\widetilde{\Phi}\left|\widetilde{\Sigma}_u, \theta\right.) = f_{MN}^{(1+np)\times n}(\widetilde{\Phi}\left|\,\hat{\widetilde{\Phi}}(\theta), \widetilde{T}\Gamma_{xx}^{*}(\theta), (\widetilde{\Sigma}_u)^{-1}\right.), \tag{20}$$

bądź w zapisie dla wielowymiarowego rozkładu normalnego (por. np. Poirier, 1995):

$$p(vec(\widetilde{\Phi})\left|\widetilde{\Sigma}_u, Y\right.) = f_N^{(1+np)n}(vec(\widetilde{\Phi})\left|vec(\hat{\widetilde{\Phi}}), \widetilde{\Sigma}_u \otimes [\widetilde{T}\Gamma_{xx}^{*}(\theta)]^{-1})\right.], \tag{21}$$

gdzie $\otimes$ to iloczyn Kroneckera, zaś *vec(.)* wektoryzację macierzy. Macierze parametrów:

$$\hat{\widetilde{\Phi}}(\theta) = [\Gamma_{xx}^{*}(\theta)]^{-1}\Gamma_{xy}^{*}(\theta) \quad \text{i} \quad \hat{\widetilde{\Sigma}}_u(\theta) = \Gamma_{yy}^{*}(\theta) - \Gamma_{yx}^{*}(\theta)[\Gamma_{xx}^{*}(\theta)]^{-1}\Gamma_{xy}^{*}(\theta), \tag{22}$$

są estymatorami największej wiarygodności i najmniejszych kwadratów macierzy współczynników $\widetilde{\Phi}$ i macierzy kowariancji $\widetilde{\Sigma}_u$. Macierz $\Gamma_{xx}^{*}(\theta)$ jest odwracalna, jeśli w modelu równowagi ogólnej liczba zakłóceń występujących w postaci strukturalnej jest równa liczbie *n* obserwowanych szeregów czasowych w równaniu obserwacji (por. np. Del Negro i Schorfheide, 2004). Macierze $\hat{\widetilde{\Phi}}(\theta)$ i $\hat{\widetilde{\Sigma}}_u(\theta)$ zywane są funkcjami ograniczeń (ang. *restriction function*), które nałożone na parametry wektorowej autoregresji bez restrykcji powodują, że naśladuje ona model równowagi ogólnej, stając się jego aproksymacją i tworząc model wektorowej autoregresji z ograniczeniami.

5. ROLA PARAMETRU WAGOWEGO

Model hybrydowej wektorowej autoregresji, w najprostszym przypadku, mógłby zostać zbudowany poprzez zastąpienie parametrów wektorowej autoregresji bez ograniczeń: Φ i Σ_u, estymatorami: $\hat{\widetilde{\Phi}}(\theta)$ i $\hat{\widetilde{\Sigma}}_u(\theta)$, będącymi funkcją parametrów θ modelu równowagi ogólnej, i zdefiniowanie dla nich rozkładu *a priori*. Byłoby to możliwe, ponieważ funkcja wiarygodności $\ell(\Phi, \Sigma_u\left|Y\right.)$, przy założeniu normalności rozkładu warunkowej gęstości obserwacji Y, zależy wyłącznie od jej parametrów Φ oraz Σ_u. Otrzymalibyśmy w ten sposób wektorową autoregresję, w której wnioskowanie *a posteriori* sprowadzałoby się do wnioskowania o parametrach modelu strukturalnego, pośrednio poprzez model autoregresji:

$$p(\theta\left|Y\right.) \propto \ell(\hat{\widetilde{\Phi}}(\theta), \hat{\widetilde{\Sigma}}_u(\theta)\left|Y\right.)\, p(\hat{\widetilde{\Phi}}(\theta), \hat{\widetilde{\Sigma}}_u(\theta)), \tag{23}$$

gdzie funkcja wiarygodności $\ell(\hat{\tilde{\Phi}}(\theta),\hat{\tilde{\Sigma}}_u(\theta)|Y)$ byłaby zależnością wyłącznie od parametrów modelu strukturalnego, zaś rozkład *a priori* $p(\hat{\tilde{\Phi}}(\theta),\hat{\tilde{\Sigma}}_u(\theta))$, byłby funkcją rozkładu *a priori* dla θ. Procedura taka oznaczałaby przyjęcie założenia, że model równowagi ogólnej jest poprawnie wyspecyfikowany a jedynie nieznane pozostają wartości parametrów θ.

Estymowane modele równowagi ogólnej zazwyczaj nie w pełni poprawnie opisują analizowane zjawisko, co powoduje nieadekwatność powyższego schematu wnioskowania i konieczność jego modyfikacji w kierunku dopuszczenia możliwości wystąpienia błędu specyfikacji. Łączny rozkład *a priori* parametrów wektorowej autoregresji $p(\Phi,\Sigma_u)$ rozważa się warunkowo względem parametrów modelu strukturalnego $p(\Phi,\Sigma_u|\theta)$, ustalając jego wartości oczekiwane w punktach $\hat{\tilde{\Phi}}(\theta)$ i $\hat{\tilde{\Sigma}}_u(\theta)$ oraz definiując dodatkowy parametr λ określający jego rozproszenie. Oznacza to, że rozkład *a priori* $p(\Phi,\Sigma_u)$, rozważony warunkowo względem θ i λ: $p(\Phi,\Sigma_u|\theta, \lambda)$, jest scentrowany w punktach odpowiadających funkcjom restrykcji z modelu równowagi ogólnej, $\hat{\tilde{\Phi}}(\theta)$ i $\hat{\tilde{\Sigma}}_u(\theta)$. Symptomem niepoprawnej specyfikacji modelu równowagi ogólnej są, w tym kontekście, odchylenia parametrów wektorowej autoregresji bez ograniczeń Φ i Σ_u od parametrów wektorowej autoregresji z restrykcjami $\hat{\tilde{\Phi}}(\theta)$ i $\hat{\tilde{\Sigma}}_u(\theta)$; niewystępowanie takich odchyleń sugeruje poprawną specyfikację modelu równowagi ogólnej.

Zdefiniowanie hiperparametru λ prowadzi, w przypadku ciągłym, do continuum, modeli hybrydowych, których własności określa zakres możliwych jego wartości. Dolna granica dla λ, powiązana z liczbą n obserwowanych szeregów czasowych i z rzędem p opóźnienia w wektorowej autoregresji, wynika z warunku istnienia rozkładu *a priori* $p(\Phi,\Sigma_u|\theta, \lambda)$ i jest ustalana na poziomie $(n+ k^*)/T$ gdzie k^* oznacza rząd macierzy $\tilde{X}$. Ogólna interpretacja wskazuje, że wartości λ bliskie dolnej granicy oznaczają, że rozkład *a priori* $p(\Phi,\Sigma_u|\theta, \lambda)$ jest znacznie rozproszony i do modelu hybrydowego jest wnoszona nikła informacja *a priori* o Φ i Σ_u, pochodząca z modelu równowagi ogólnej, co oznacza że posiada on własności bliskie wektorowej autoregresji bez restrykcji. Z drugiej strony, graniczny przypadek, kiedy $\lambda\to\infty$, wskazuje, że rozkład $p(\Phi,\Sigma_u|\theta, \lambda)$ zmierza do rozkładu jednopunktowego w modalnej $\hat{\tilde{\Phi}}(\theta)$ i $\hat{\tilde{\Sigma}}_u(\theta)$, a model hybrydowy naśladuje model równowagi ogólnej.

Zdefiniowanie parametru określającego rozproszenie rozkładu *a priori* parametrów wektorowej autoregresji wskazuje, że warunkowo względem λ, wnioskowanie *a posteriori* w modelu hybrydowym przebiega w sposób standardowy. Łączny rozkład *a priori* parametrów modelu równowagi ogólnej i wektorowej autoregresji ma strukturę hierarchiczną:

$$p(\Phi,\Sigma_u,\theta|\lambda) = p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)\,, \tag{24}$$

i po uwzględnieniu warunkowej gęstości obserwacji, $p(Y|\Phi,\Sigma_u)$, prowadzi do łącznego rozkładu obserwacji i parametrów modelu hybrydowego, warunkowo względem λ:

$$p(Y,\Phi,\Sigma_u,\theta|\lambda) = p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta), \tag{25}$$

gdzie warunkowa gęstość obserwacji: $p(Y|\Phi,\Sigma_u)$ nie zależy od θ i λ. Łączny rozkład *a posteriori*, przy ustalonym λ, jest dany przez:

$$p(\Phi,\Sigma_u,\theta|Y,\lambda) = \frac{p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)}{p(Y|\lambda)}\,. \tag{26}$$

Na podstawie jego dekompozycji otrzymujemy warunkowy rozkład *a posteriori* współczynników hybrydowej autoregresji i brzegowy dla θ:

$$p(\Phi,\Sigma_u,\theta|Y,\lambda) = p(\Phi,\Sigma_u|Y,\theta,\lambda)\,p(\theta|Y,\lambda)\,, \tag{27}$$

gdzie warunkowy rozkład *a posteriori* parametrów autoregresji $p(\Phi,\Sigma_u|Y,\theta,\lambda)$ jest wyznaczany ze wzoru Bayesa, na podstawie formalnego wnioskowania w modelu wektorowej autoregresji, warunkowo względem parametrów modelu równowagi ogólnej i λ:

$$p(\Phi,\Sigma_u|Y,\theta,\lambda) = \frac{p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)}{\int p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,d(\Phi,\Sigma_u)}\,. \tag{28}$$

Optymalna wartości parametru rozproszenia λ jest ustalana na podstawie kryterium maksymalizacji brzegowej gęstości obserwacji:

$$\begin{aligned} p(Y|\lambda) &= \int\int p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)\,d(\Phi,\Sigma_u)\,d\theta = \\ &= \int p(\theta)\int p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,d(\Phi,\Sigma_u)\,d\theta = \int p(\theta)\,p(Y|\theta,\lambda)\,d\theta \end{aligned} \tag{29}$$

traktowanej jako ogólna miara dopasowania modelu do danych. Całkowania w funkcji brzegowej gęstości obserwacji $p(Y|\lambda)$ względem Φ i Σ_u, przy ustalonym λ i wektorze θ, mogą zostać wykonane analitycznie, jeśli rozkład *a priori* $p(\Phi,\Sigma_u\theta, \lambda)$ przyjmie się z naturalnej rodziny sprzężonej do rozkładu normalnego dla $p(Y|\Phi,\Sigma_u)$, natomiast jedynie ostatnia całka, po elementach wektora θ, jest obliczana numerycznie.

W modelu hybrydowym możliwe jest potraktowanie λ jako dodatkowego parametru, podlegającego estymacji po przyjęciu dla niego rozkładu *a priori*, w szczególności rozkładu jednostajnego na ustalonym arbitralnie przedziale, (zob. Adjemian i in., 2008). Łączny rozkład *a priori* wszystkich parametrów modelu hybrydowego ma wtedy postać:

$$p(\Phi,\Sigma_u,\theta,\lambda) = p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)\,p(\lambda)\,, \tag{30}$$

gdzie zmienne losowe θ i λ są niezależne. Łączny rozkład obserwacji, parametrów modelu hybrydowego i parametru wagowego jest wtedy określony przez:

$$p(Y,\Phi,\Sigma_u,\theta,\lambda) = p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)\,p(\lambda)\,, \tag{31}$$

z którego uzyskujemy łączny rozkład *a posteriori*:

$$p(\Phi,\Sigma_u,\theta,\lambda|Y) = \frac{p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)\,p(\lambda)}{p(Y)}, \tag{32}$$

gdzie *p(Y)* jest brzegową gęstością obserwacji, wyrażoną całką:

$$\begin{aligned} p(Y) &= \int\int\int p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)\,p(\theta)\,p(\lambda)d(\Phi,\Sigma_u)d\theta d\lambda = \\ &= \int p(\lambda)\int p(\theta)\int p(Y|\Phi,\Sigma_u)\,p(\Phi,\Sigma_u|\theta,\lambda)d(\Phi,\Sigma_u)\,d\theta d\lambda = \\ &= \int p(\lambda)\int p(\theta)\,p(Y|\theta,\lambda)d\theta d\lambda\,, \end{aligned} \tag{33}$$

natomiast brzegowy rozkład *a posteriori* dla λ jest otrzymywany z łącznego rozkładu *a posteriori* po zapisaniu:

$$\begin{aligned} p(\lambda|Y) &= \int p(\Phi,\Sigma_u,\theta,\lambda|Y)d(\Phi,\Sigma_u,\theta) = \\ &= \int p(\lambda|Y,\Phi,\Sigma_u,\theta)p(\Phi,\Sigma_u|Y,\theta)p(\theta|Y)d(\Phi,\Sigma_u,\theta). \end{aligned} \tag{34}$$

Łączny rozkład *a posteriori* można poddać dekompozycji na warunkowy rozkład *a posteriori* współczynników hybrydowej autoregresji i brzegowy dla θ i λ:

$$p(\Phi,\Sigma_u,\theta,\lambda|Y) = p(\Phi,\Sigma_u|Y,\theta,\lambda)p(\theta,\lambda|Y), \tag{35}$$

gdzie warunkowy rozkład *a posteriori* parametrów autoregresji $p(\Phi,\Sigma_u|Y,\theta,\lambda)$ wyraża się standardową gęstością prawdopodobieństwa, natomiast brzegowy rozkład parametrów modelu strukturalnego i parametru wagowego $p(\theta,\lambda|Y)$, jest przybliżany numerycznie, z zastosowaniem algorytmu Metropolisa i Hastingsa (zob. Adjemian i in., 2008).

Przypadek $\lambda=0$ implikuje, że w modelu hybrydowym parametry modelu równowagi ogólnej i wektorowej autoregresji są *a priori* niezależne:

$$p(\Phi,\Sigma_u,\theta|\lambda) = p(\Phi,\Sigma_u|\theta,\lambda)p(\theta) = p(\Phi,\Sigma_u)p(\theta), \tag{36}$$

skąd łączny rozkład *a posteriori* ma postać: $p(Y,\Phi,\Sigma_u,\theta) = \ell(\Phi,\Sigma_u|Y)p(\Phi,\Sigma_u)p(\theta)$.

Funkcja wiarygodności $\ell(\Phi,\Sigma_u|Y)$ nie zależy od wektora parametrów modelu strukturalnego, co oznacza że rozkład *a priori* $p(\theta)$ nie jest uaktualniany przez obserwacje i dane empiryczne nie dostarczają informacji o θ. W konsekwencji rozkład *a posteriori* parametrów modelu równowagi ogólnej, uzyskany na podstawie modelu hybrydowego, jest tożsamy z rozkładem *a priori*. Przyjęcie $\lambda>0$ implikuje zależność *a priori* Φ, Σ_u i θ, umożliwiając jednocześnie uwzględnienie informacji z modelu strukturalnego w modelu hybrydowym oraz wnioskowanie na jego podstawie o θ. W modelu hybrydowym ograniczenie nakładane na parametr wagowy: $\lambda>(n+k^*)/T$, oznacza że w granicznym przypadku wiedza wstępna pochodząca z modelu równowagi ogólnej jest uwzględniana w znikomym stopniu, jednak wystarczającym do przeprowadzenia formalnego wnioskowania o parametrach modelu równowagi ogólnej.

6. KONSTRUKCJA ROZKŁADU *A PRIORI*

Specyfikacja warunkowego, względem parametrów modelu strukturalnego i przy ustalonym parametrze λ, rozkładu *a priori* $p(\Phi,\Sigma_u|\theta, \lambda)$ współczynników wektorowej autoregresji bez restrykcji, wykorzystuje formalne wnioskowanie statystyczne, które wywodzi się z koncepcji dolosowania danych. Rozważamy pomocniczy model wektorowej autoregresji: $Y^* = X^*\Phi + U^*$, gdzie Y^* jest macierzą $(T^*\times n)$, X^* to macierz $(T^*\times(1+np))$, U^* jest macierzą $(T^*\times n)$ o wierszach $u_t^{*\prime}$, $u_t^* \sim iidN^{(n)}(0,\Sigma_u)$ rozważamy jego estymację na podstawie próbki danych symulacyjnych, pochodzących z liniowego rozwiązania modelu równowagi ogólnej. Skalar λ, opisujący rozproszenie rozkładu *a priori* z modelu strukturalnego, określa liczebność danych sztucznych T^* względem danych rzeczywistych, $\lambda=T^*/T$, która odzwierciedla wagę poszczególnych modeli we wnioskowaniu na podstawie modelu hybrydowego $W = T^*/(T + T^*) = \lambda/(1 + \lambda)$. Łączny rozkład *a posteriori* współczynników i macierzy kowariancji wektorowej autoregresji szacowanej na danych sztucznych, przy założeniu dla nich *a priori* rozkładu Jeffreysa: $p(\Phi,\Sigma_u) \propto \det(\Sigma_u)^{-(n+1)/2}$, ma postać rozkładu macierzowego normalnego – odwróconego Wisharta:

$$p(\Phi,\Sigma_u\left|Y^*,\lambda\right.) = f_{MNIW}(\Phi,\Sigma_u\left|\ \hat{\Phi}^*,(X^{*\prime}X^*),(\hat{\Sigma}_u^*)^{-1},\lambda T-(1+np)\right.), \tag{37}$$

gdzie $\hat{\Phi}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$ i $\hat{\Sigma}_u^* = Y^{*\prime}Y^* - Y^{*\prime}X^*(X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$, (por. np. Zellner, 1971; Poirier, 1995; Koop, 2003). Rozkład $p(\Phi,\Sigma_u|Y^*,\lambda)$ aproksymujący $p(\Phi,\Sigma_u|\theta, \lambda)$, jest przyjmowany jako rozkład *a priori* dla modelu wektorowej autoregresji, estymowanego na danych rzeczywistych. Rozkład ten należy do naturalnej rodziny sprzężonej do rozkładu normalnego dla funkcji wiarygodności modelu autoregresji dla danych rzeczywistych $\ell(\Phi,\Sigma_u\left|Y\right.)$, w szczególności jest to iloczyn rozkładu odwróconego Wisharta dla Σ_u i macierzowego normalnego dla Φ, warunkowo względem Σ_u.

Znaczna liczebność danych sztucznych względem liczebności danych obserwowanych oznacza, że estymatory parametrów modelu hybrydowego będą pod silnym wpływem informacji wywodzących się z modelu równowagi ogólnej, a dokładniej z aproksymacji wektorową autoregresją rozwiązania zlinearyzowanej postaci modelu strukturalnego. Jeśli pomijamy dane symulacyjne w procesie wnioskowania, $(\lambda=T^*=0)$, to oszacowania parametrów modelu hybrydowego sprowadzają się do ocen parametrów uzyskanych w wektorowej autoregresji z nieinformacyjnym rozkładem *a priori*, zaś model równowagi ogólnej nie przekazuje informacji wynikającej z przyjętych do jego konstrukcji założeń. Optymalna wartość λ, znajdująca się w górnym zakresie możliwych jej wartości, w szczególności $(\lambda\to\infty)$, oznacza że model hybrydowy ma własności bliskie modelowi strukturalnemu, co bywa również uważane za argument przemawiający za odpowiednią jego specyfikacją, spełnieniem przez obserwacje restrykcji ekonomicznych i dobrym dopasowaniem do danych empirycznych.

Stosowanie procedury dolosowania danych sztucznych, jako metody wprowadzania do modelu hybrydowego informacji wstępnych, jest związane z pojawieniem się dodatkowej niepewności próbkowej, będącej konsekwencją symulacji obserwacji. Możliwa jest, w takiej sytuacji, modyfikacja wnioskowania i zastąpienie momentów empirycznych danych generowanych ich momentami teoretycznymi. W przypadku estymowanych modeli równowagi ogólnej można analityczne wyznaczyć teoretyczne momenty drugiego rzędu zmiennych obserwowalnych, na podstawie reprezentacji modelu równowagi ogólnej w przestrzeni stanów, i zastąpić nimi występujące w funkcji wiarygodności wektorowej autoregresji momenty empiryczne danych sztucznych (por. np. Del Negro i Schorfheide, 2004; Del Negro i Schorfheide, 2008). Rozwiązanie modelu strukturalnego w formie równania stanu jest dyskretnym liniowym układem dynamicznym, rzędu pierwszego, którego wartość oczekiwana oraz momenty drugiego rzędu, dla ustalonego θ, mogą zostać wyznaczone analitycznie.

Wektor wartości oczekiwanej zmiennych endogenicznych modelu równowagi ogólnej w stanie stabilnym $E(Y_t)$ jest równy F. Scentrowana macierz kowariancji wektora Y_t, wyznaczona na podstawie reprezentacji modelu strukturalnego w przestrzeni stanów, jest określona przez $\Sigma_y^{(0)} = C'\Sigma_s C + \Sigma_v$, natomiast pozostałe macierze (scentrowanych) autokowariancji są wyrażone formułą, $\Sigma_y^{(j)} = C' A^j \Sigma_s C$, dla $j=1,2,...p$, oraz $\Sigma_y^{(j)} = \Sigma_y^{(-j)}$, (zob. Warne i in., 2012). Macierz kowariancji zmiennych stanu w stanie stabilnym Σ_s jest bezpośrednio obliczana po rozwiązaniu równania Lapunowa:

$$A\Sigma_s A' + B\Sigma_\varepsilon B' = \Sigma_s , \tag{38}$$

które jest otrzymywane na podstawie definicji macierzy kowariancji wektora stanu:

$$E[(s_t - \bar{s})(s_t - \bar{s})'] = E[(As_{t-1} + B\varepsilon_t)(As_{t-1} + B\varepsilon_t)'] = A\Sigma_{s,(t)}A' + B\Sigma_{\varepsilon,(t)}B' = \Sigma_{s,(t)},$$

przy założeniu niezmienności w czasie $\Sigma_{s,(t)} = \Sigma_s$ i macierzy kowariancji szoków strukturalnych $\Sigma_{\varepsilon,(t)} = \Sigma_\varepsilon$, spełnienia przez macierz A warunków stabilności, niezależności wektora stanu s_t od wektora zakłóceń strukturalnych ε_t, oraz $E(s_t) = \bar{s} = 0$.

Aproksymacja modelu równowagi ogólnej modelem wektorowej autoregresji bez restrykcji jest dokonywana poprzez zastąpienie jej współczynników $\widetilde{\Phi}$ i $\widetilde{\Sigma}_u$ parametrami implikowanymi przez model strukturalny $\widetilde{\Phi}(\theta)$ i $\widetilde{\Sigma}_u(\theta)$, uzyskanymi na podstawie autoregresji dla momentów teoretycznych, warunkowo względem θ. Wartość oczekiwana wektora Y_t jest równa $E(Y_t) = F$ w modelu strukturalnym i $E(\widetilde{Y}_t) = \Phi'(\theta)\widetilde{X}_t$ w wektorowej autoregresji. Macierze niecentralnych momentów drugiego rzędu zmiennych obserwowalnych $\widetilde{Y}_t$ i $\widetilde{X}_t$, względem rozkładu parametrów modelu strukturalnego θ, są dane przez wartości oczekiwane (zob. Warne i in., 2012):

$$E_\theta(\widetilde{Y}_t\ \widetilde{Y}_t') = E_\theta[(\widetilde{\Phi}'\widetilde{X}_t + \widetilde{u}_t)(\widetilde{\Phi}'\widetilde{X}_t + \widetilde{u}_t)'] = E_\theta[(\widetilde{\Phi}'\widetilde{X}_t(\widetilde{\Phi}'\widetilde{X}_t)'] + E_\theta(\widetilde{u}_t\widetilde{u}_t') = \\ = \widetilde{\Phi}'X_tX_t\widetilde{\Phi} + \widetilde{\Sigma}_u = FF' + \Sigma_y^{(0)} = \Gamma_{yy}^*(\theta), \tag{39}$$

gdzie $E(Y_tY_t') = FF' + \Sigma_y^{(0)}$ w modelu równowagi ogólnej,

$$E_\theta(\widetilde{X}_t\ \widetilde{X}_t') = E\begin{bmatrix} 1 & \widetilde{Y}_{t-1}' & \dots & \widetilde{Y}_{t-p}' \\ \widetilde{Y}_{t-1} & \widetilde{Y}_{t-1}\widetilde{Y}_{t-1}' & \dots & \widetilde{Y}_{t-1}\widetilde{Y}_{t-p}' \\ \vdots & \vdots & \vdots & \vdots \\ \widetilde{Y}_{t-p} & \widetilde{Y}_{t-p}\widetilde{Y}_{t-1}' & \dots & \widetilde{Y}_{t-p}\widetilde{Y}_{t-p}' \end{bmatrix} = E_\theta\begin{bmatrix} 1 & F & \dots & F \\ F & \Sigma_y^{(0)} & \dots & \Sigma_y^{p-1} \\ \vdots & \vdots & \vdots & \vdots \\ F & \Sigma_y^{p-1} & \dots & \Sigma_y^{(0)} \end{bmatrix} = \Gamma_{xx}^*(\theta), \tag{40}$$

dla $E_\theta(\widetilde{Y}_t\widetilde{Y}_{t-j}') = E_\theta[(\widetilde{\Phi}'\widetilde{X}_t + \widetilde{u}_t)(\widetilde{\Phi}'\widetilde{X}_{t-j} + \widetilde{u}_{t-j})'] = FF' + \Sigma_y^{(j)}$, $E_\theta(\widetilde{Y}_t\widetilde{Y}_{t-j}') = E_\theta(\widetilde{Y}_{t-j}\widetilde{Y}_t')$ oraz:

$$E_\theta(\widetilde{X}_t\ \widetilde{Y}_t') = E_\theta\begin{bmatrix} \widetilde{Y}_t' \\ \widetilde{Y}_{t-1}\widetilde{Y}_t' \\ \vdots \\ \widetilde{Y}_{t-p}\widetilde{Y}_t' \end{bmatrix} = E_\theta\begin{bmatrix} F' \\ FF' + \Sigma_y^{(1)} \\ \vdots \\ FF' + \Sigma_y^{(p)} \end{bmatrix} = \Gamma_{xy}^*(\theta), \tag{41}$$

przy czym $E_\theta(\widetilde{X}_t\ \widetilde{Y}_t') = E_\theta(\widetilde{Y}_t\ \widetilde{X}_t') = \Gamma_{yx}^*(\theta)$. Model wektorowej autoregresji $\widetilde{Y}_t$ względem $\widetilde{X}_t$, bazujący na momentach teoretycznych determinuje rzutowanie parametrów modelu równowagi ogólnej θ w parametry wektorowej autoregresji, $\hat{\widetilde{\Phi}}(\theta)$ i $\hat{\widetilde{\Sigma}}_u(\theta)$.

Zamiana momentów empirycznych danych sztucznych $Y^{*\prime}Y^*$, $Y^{*\prime}X^*$ i $X^{*\prime}X^*$ na momenty teoretyczne $\Gamma_{yy}^*(\theta)$, $\Gamma_{xy}^*(\theta)$ i $\Gamma_{xy}^*(\theta)$, pochodzące z modelu równowagi ogólnej, prowadzi do następującego łącznego rozkładu *a priori* dla parametrów hybrydowego modelu wektorowej autoregresji:

$$p(\Phi, \Sigma_u|\theta, \lambda) = c_{(1)}^{-1}(\theta)\det(\Sigma_u)^{-(\lambda T+n+1)/2} \times \\ \exp\{-0.5tr[\lambda T\Sigma_u^{-1}(\Gamma_{yy}^*(\theta) - \Phi'\Gamma_{xy}^*(\theta) - \Gamma_{yx}^*(\theta)\Phi + \Phi'\Gamma_{xx}^*(\theta)\Phi)]\}, \tag{42}$$

gdzie $\widetilde{T} = \lambda T$, zaś stała normująca z rozkładu Wisharta jest równa:

$$c_{(1)}(\theta)=(2\pi)^{\frac{n(1+np)}{2}}\det(\lambda T\Gamma^{*}_{xx}(\theta))^{-\frac{n}{2}}\det(\lambda T\Sigma_u(\theta))^{-\frac{\lambda T-(1+np)}{2}}2^{\frac{n(\lambda T-(1+np))}{2}}\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma\left[\frac{\lambda T-(1+np)+1-i}{2}\right]. \quad (43)$$

Spełnienie warunku: $\lambda \geq (1+np+n)/T$ oraz odwracalność macierzy momentów $\Gamma^{*}_{xx}(\theta)$, oznaczają, że otrzymany rozkład jest właściwy (funkcja gęstości całkuje się do jedności) oraz niezdegenerowany (jego dziedzina nie jest ograniczona do podprzestrzeni przestrzeni parametrów wektorowej autoregresji). Jeśli $\lambda=0$ to model połączony sprowadza się do wektorowej autoregresji bez restrykcji, natomiast jeśli $0<\lambda<(1+np+n)/T$ to otrzymany rozkład *a priori* jest niewłaściwy.

7. ROZKŁAD A POSTERIORI

Łączny rozkład *a posteriori* dla Φ i Σ_u, proporcjonalny do iloczynu funkcji wiarygodności $\ell(\Phi,\Sigma_u|Y)$ i rozkładu *a priori* $p(\Phi,\Sigma_u|\theta,\lambda)$:

$$p(\Phi,\Sigma_u|Y,\theta,\lambda)\propto \ell(\Phi,\Sigma_u|Y)\,p(\Phi,\Sigma_u|\theta,\lambda), \quad (44)$$

ma postać rozkładu macierzowego normalnego – odwróconego Wisharta:

$$p(\Phi,\Sigma_u|Y,\theta,\lambda)=f_{MNIW}(\Phi,\Sigma_u|\ \hat{\Phi}^{m},[\lambda T\Gamma^{*}_{xx}(\theta)+X'X],[(\lambda+1)T\hat{\Sigma}^{m}_{u}(\theta)]^{-1},(1+\lambda)T-(1+np)), \quad (45)$$

i dany jest wzorem:

$$\begin{aligned}p(\Phi,\Sigma_u|Y,\theta,\lambda)=c^{-1}_{(2)}(\theta)\det(\Sigma_u)^{-[(1+\lambda)T+n+1]/2}\exp\{-0.5tr\{\Sigma_u^{-1}[(Y-X\Phi)'(Y-X\Phi)+\\ +\lambda T(\Gamma^{*}_{yy}(\theta)-\Phi'\Gamma^{*}_{xy}(\theta)-\Gamma^{*}_{yx}(\theta)\Phi+\Phi'\Gamma^{*}_{xx}(\theta)\Phi)]\}\},\end{aligned} \quad (46)$$

gdzie stała normująca jest równa:

$$\begin{aligned}c_{(2)}(\theta)=(2\pi)^{\frac{n(1+np)}{2}}\det(\lambda T\Gamma^{*}_{xx}(\theta)+X'X)^{-\frac{n}{2}}\det((1+\lambda)T\,\Sigma^{m}_{u}(\theta))^{-\frac{(1+\lambda)T-(1+np)}{2}}\times\\ \times 2^{\frac{n((1+\lambda)T-(1+np))}{2}}\pi^{\frac{n(n-1)}{4}}\prod_{i=1}^{n}\Gamma\left[\frac{\lambda T-(1+np)+1-i}{2}\right].\end{aligned} \quad (47)$$

Układ warunkowych rozkładów *a posteriori* ma postać:

$$p(\Sigma_u|Y,\theta,\lambda)=f_{IW}(\Sigma_u|\ (\lambda+1)T\hat{\Sigma}^{m}_{u}(\theta),(1+\lambda)T-(1+np),n), \quad (48)$$

$$p(\Phi|Y,\Sigma_u,\theta,\lambda) = f_{MN}^{(1+np)\times n}(\Phi|\,\hat{\Phi}^m(\theta),(\lambda T\Gamma_{xx}^*(\theta)+X'X),\Sigma_u^{-1}), \tag{49}$$

gdzie wartość oczekiwana $\hat{\Phi}^m(\theta) = (\lambda T\Gamma_{xx}^*(\theta)+X'X)^{-1}(\lambda T\Gamma_{xy}^*(\theta)+X'Y)$ współczynników wektorowej autoregresji jest wypadkową momentów teoretycznych, pochodzących z modelu równowagi ogólnej, i momentów empirycznych danych obserwowalnych, oraz:

$$\hat{\Sigma}_u^m(\theta) = [(\lambda+1)T]^{-1}[(\lambda T\Gamma_{yy}^*(\theta)+Y'Y)-(\lambda T\Gamma_{yx}^*(\theta)+Y'X)\hat{\Phi}^m(\theta)]. \tag{50}$$

Macierze momentów teoretycznych, występujące w konstrukcji rozkładu *a priori* na podstawie modelu strukturalnego, mają również uzasadnienie, jako macierze graniczne, przy liczbie danych sztucznych T^* zmierzającej do nieskończoności: $\lim_{T^*\to\infty} Y^{*\prime}Y^*/T^* = \Gamma_{yy}^*(\theta)$. Analogiczne zależności zachodzą dla $\Gamma_{xx}^*(\theta)$ i $\Gamma_{xy}^*(\theta)$. Liczba obserwacji sztucznych $T^*\to\infty$, w modelu połączonym oznaczałaby, że wnioskowanie *a posteriori* sprowadza się do wnioskowania na podstawie modelu strukturalnego. Analiza zależności wartości oczekiwanej *a posteriori* od parametru wagowego λ jest możliwa po wyłączeniu wyrażenia $T(1+\lambda)$ z każdego z czynników występujących w równaniu dla $\hat{\Phi}^m(\theta)$:

$$\hat{\Phi}^m(\theta) = \Big(\frac{\lambda}{1+\lambda}\Gamma_{xx}^*(\theta)+\frac{1}{1+\lambda}T^{-1}X'X\Big)^{-1}\Big(\frac{\lambda}{1+\lambda}\Gamma_{xy}^*(\theta)+\frac{1}{1+\lambda}T^{-1}X'Y\Big), \tag{51}$$

jeśli $\lambda\to\infty$ to $\hat{\Phi}^m(\theta)$ zmierza do funkcji ograniczenia $\tilde{\Phi}(\theta)$, natomiast jeśli $\lambda\to 0$, to $\hat{\Phi}^m(\theta)$ jest bliskie estymatorowi najmniejszych kwadratów dla Φ, przy czym $\lambda\geq(1+np+n)/T$.

Konstrukcja modelu hybrydowego umożliwia wnioskowanie *a posteriori* o parametrach modelu równowagi ogólnej, które przebiega pośrednio, poprzez wnioskowanie o parametrach wektorowej autoregresji Φ i Σ_u, warunkowo względem współczynnika wagowego λ. Brzegowy rozkład *a posteriori* $p(\theta|Y,\lambda)$, jest uzyskiwany na podstawie wzoru Bayesa: $p(\theta|Y,\lambda) = p(Y|\theta,\lambda)p(\theta)/\,p(Y|\lambda)$, gdzie $p(Y|\lambda) = \int p(Y|\theta,\lambda)p(\theta)d\theta$ jest brzegową, względem wszystkich parametrów modelu hybrydowego, gęstością obserwacji. Warunkowa względem θ i λ oraz brzegowa względem Φ i Σ_u gęstość obserwacji w modelu hybrydowym dana jest przez:

$$p(Y|\theta,\lambda)=\int p(Y|\Phi,\Sigma_u)p(\Phi,\Sigma_u|\theta,\lambda)d(\Phi,\Sigma_u)=\frac{p(Y|\Phi,\Sigma_u)p(\Phi,\Sigma_u|\theta,\lambda)}{p(\Phi,\Sigma_u|Y,\theta,\lambda)}=$$

$$=(2\pi)^{-\frac{nT}{2}}\frac{c_{(2)}(\theta)}{c_{(1)}(\theta)}=\frac{\det(\lambda T\Gamma^*_{xx}(\theta)+X'X)^{-\frac{n}{2}}\det((1+\lambda)T\hat{\Sigma}^m_u(\theta))^{-\frac{(1+\lambda)T-(1+np)}{2}}}{\det(\lambda T\Gamma^*_{xx}(\theta))^{-\frac{n}{2}}\det(\lambda T\hat{\hat{\Sigma}}_u(\theta))^{-\frac{\lambda T-(1+np)}{2}}}\times \tag{52}$$

$$\times\frac{(2\pi)^{-\frac{nT}{2}}2^{\frac{n((1+\lambda)T-(1+np))}{2}}\prod_{i=1}^{n}\Gamma\left[\frac{(1+\lambda)T-(1+np)+1-i}{2}\right]}{2^{\frac{n(\lambda T-(1+np))}{2}}\prod_{i=1}^{n}\Gamma\left[\frac{\lambda T-(1+np)+1-i}{2}\right]}.$$

Charakterystyki rozkładu *a posteriori:* $p(\theta|Y,\lambda)\propto\ell(\theta|Y,\lambda)p(\theta)$ są otrzymywane numerycznie, z zastosowaniem algorytmu Metropolisa i Hastingsa (zob. np. Schorfheide, 2000; Del Negro i Schorfheide, 2004; Adjemian i in., 2008). W przypadku, kiedy λ traktujemy jako dodatkowy parametr podlegający estymacji, łączny rozkład *a posteriori* $p(\Phi,\Sigma_u,\theta,\lambda|Y)$ jest iloczynem rozkładu macierzowego normalnego – odwróconego Wisharta dla $p(\Phi,\Sigma_u|Y,\theta,\lambda)$ i rozkładu prawdopodobieństwa o niestandardowej postaci dla $p(\theta,\lambda|Y)$, przybliżanej przez MCMC.

8. PODSUMOWANIE

Opracowanie zawiera omówienie budowy i wnioskowania w modelu wektorowej autoregresji, w którym rozkład *a priori* dla macierzy współczynników i kowariancji jest generowany na podstawie estymowanego modelu równowagi ogólnej. Przyjęcie do konstrukcji rozkładu *a priori* modelu ekonometrycznego, u podstaw którego leżą założenia implikowane przez teorię mikro i makroekonomii, pozwala na uwzględnienie różnych jej aspektów w modelu połączonym, w szczególności umożliwia testowanie stopnia potwierdzenia alternatywnych założeń przez dane empiryczne. Kluczową rolę w modelu połączonym wektorowej autoregresji i równowagi ogólnej odgrywa parametr wagowy, odpowiedzialny za udział informacji *a priori*. Parametry modelu połączonego są estymowane technikami bayesowskimi, co umożliwia formalne ujęcie niepewności związanej z ocenami parametrów *a posteriori*. Artykuł stanowił syntezę informacji teoretycznych związanych z metodologią DSGE-VAR, będących etapem wstępnym badań empirycznych.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

[1] Adjemian A., DarracqPariès M., Moyen S., (2008), Towards a Monetary Policy Evaluation Framework, *European Central Bank Working Paper* No. 942, Frankfurt am Main, Germany.
[2] Adolfson M., Laseén S., Lindé J., Villani M., (2008), Evaluating an Estimated New Keynesian Small Open Economy Model, *Journal of Economic Dynamics and Control*, Elsevier, 32 (8), 2690–2721.
[3] An S., Schorfheide F., (2007), Bayesian Analysis of DSGE Models – Rejoinder, *Econometric Reviews*, Taylor and Francis Journal, 26 (2–4), 211–219.
[4] Brzoza-Brzezina M., Kolasa M., (2012), Bayesian Evaluation of DSGE Models with Financial Frictions, *National Bank of Poland Working Paper* 109.
[5] Chow H. K., McNelis P. D., (2010), Need Singapore Fear Floating? A DSGE-VAR Approach, *Research Collection School of Economics*, Paper 1250.
[6] DeJong D. N., Ingram B. F., Whiteman C. H., (1996), A Bayesian Approach to Calibration, *Journal of Business & Economic Statistics*, American Statistical Association, 14 (1), 1–9.
[7] Del Negro M., Schorfheide F., (2004), Priors from General Equilibrium Models for VARs, *International Economic Review*, 45 (2), 643–673.
[8] Del Negro M., Schorfheide F., (2008), Forming Priors for DSGE Models, (and How It Affects the Assessment of Nominal Rigidities), *Journal of Monetary Economics*, Elsevier, 55 (7), 1191–1208.
[9] Del Negro M., Schorfheide F., Smets F., Wouters R., (2007), On the Fit of New-Keynesian Models, *Journal of Business & Economic Statistics*, American Statistical Association, 25 (2), 123–143.
[10] Doan T., Litterman R., Sims C., (1983), Forecasting and Conditional Projections Using Realistic Prior Distributions, *NBER Working Papers* 1202, National Bureau of Economic Research, Inc.
[11] Gallant R. A., McCulloch R. E., (2009), On the Determination of General Scientific Models with Application to Asset Pricing, *Journal of the American Statistical Association*, American Statistical Association, 104 (485), 117–131.
[12] Ingram B. F., Whiteman C. H., (1994), Supplanting Minnesota Prior. Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors, *Journal of Monetary Economics*, Elsevier, 34 (3), 497–510.
[13] Ireland P. N., (2004), A Method for Taking Models to the Data, *Journal of Economics Dynamic & Control*, Elsevier, 28 (6), 1205–1226.
[14] Kolasa M., Rubaszek M., Skrzypczyński P., (2012), Putting the New Keynesian DSGE Model to the Real-time Forecasting Test, *Journal of Money, Credit and Banking*, Blackwell Publishing, 44 (7), 1301–1324.
[15] Koop G., (2003), *Bayesian Econometrics*, Wiley, England.
[16] Lee K., Matheson T., Smith C., (2007), Open Economy DSGE-VAR Forecasting and Policy Analysis, Head to Head with the RBNZ Published Forecasts, *Reserved Bank of New Zealand Discussion Paper* Series, DP2007/01.
[17] Litterman R., (1986), Forecasting with Bayesian Vector Autoregression, Five Years of Experience, *Journal of Business & Economic Statistics*, American Statistical Association, 4 (1), 25–38.
[18] Liu G., Gupta R., (2008), Forecasting the South African Economy, A DSGE-VAR Approach, *Working Paper* 2008–32, Tilburg University, Center for Economic Research.
[19] Poirier D. J., (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Hong Kong.
[20] Schorfheide F., (2000), Loss Function Based Evaluation of DSGE Models, *Journal of Applied Econometrics*, John Wiley and Sons, Ltd., 15 (6), 645–670.
[21] Sims C. A., Zha T., (1998), Bayesian Methods for Dynamic Multivariate Models, *International Economic Review*, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, 39 (4), 949–968.

[22] Theil H., Goldberger A. S., (1961), On Pure and Mixed Estimation in Economics. *International Economic Review*, 2 (1), 65–78.
[23] Warne A., Coenen G., Christoffel K., (2012), Forecasting with DSGE-VAR Models, manuscript – Directorate General Research, European Central Bank, *manuscript* – Directorate General Research, European Central Bank.
[24] Watanabe T., (2007), The Application of DSGE-VAR Model to Macroeconomic Data in Japan, *ESRI Discussion Paper* Series 225–E.
[25] Wróbel-Rotter R., (2007a), Dynamic Stochastic General Equilibrium Models: Structure and Estimation, w: Welfe W., Wdowiński P. (red.), *Modelling Economies in Transition* 2006, Łódź, Wydawnictwo "Green", 9–26.
[26] Wróbel-Rotter R., (2007b), Dynamiczne Stochastyczne Modele Równowagi Ogólnej: zarys metodologii badań empirycznych, *Folia Oeconomica Cracoviensia,* 48, 69–93.
[27] Wróbel-Rotter R., (2007c), Dynamiczny Stochastyczny Model Równowagi Ogólnej: przykład dla gospodarki polskiej, *Przegląd Statystyczny,*54 (3), 25–48.
[28] Wróbel-Rotter R., (2008), Bayesian Estimation of a Dynamic General Equilibrium Model, w: Welfe A., (red.), *Metody Ilościowe w Naukach Ekonomicznych*, Ósme Warsztaty Doktorskie z zakresu Ekonometrii i Statystyki, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, 279–294.
[29] Wróbel-Rotter R., (2011a), Empiryczne modele równowagi ogólnej: gospodarstwa domowe i producent finalny, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, seria Ekonomia, 869, 109–128.
[30] Wróbel-Rotter R., (2011b), Obszary stabilności rozwiązania empirycznych modeli równowagi ogólnej: zastosowanie metod analizy wrażliwości, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, seria Metody analizy danych, 873, 121–135.
[31] Wróbel-Rotter R., (2011c), Sektor producentów pośrednich w empirycznym modelu równowagi ogólnej, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, seria Ekonomia, 872, 73–93.
[32] Wróbel-Rotter R., (2012a), Empiryczne modele równowagi ogólnej: zagadnienia numeryczne estymacji bayesowskiej, w: Sokołowski A., (red.), *Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie*, seria: Metody analizy danych, 878, 143–162.
[33] Wróbel-Rotter R., (2012b), Struktura empirycznego modelu równowagi ogólnej dla niejednorodnych gospodarstw domowych, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie,* seria Ekonomia, 879, 107–126.
[34] Wróbel-Rotter R., (2012c), Wybrane zagadnienia współczesnego modelowania strukturalnego, część I: estymowane modele równowagi ogólnej w zarysie, *Folia Oeconomica Cracoviensia,* 53, 59–83.
[35] Wróbel-Rotter R., (2012d), Wybrane zagadnienia współczesnego modelowania strukturalnego, część II: wnioskowanie w estymowanych modelach równowagi ogólnej, *Folia Oeconomica Cracoviensia*,53, 85–112.
[36] Wróbel-Rotter R., (2013a), Estymowane modele równowagi ogólnej: zastosowanie metody dekompozycji funkcji do oceny zależności między parametrami postaci strukturalnej i zredukowanej, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, seria Metody Analizy Danych, (w druku).
[37] Wróbel-Rotter R., (2013b), Estymowane modele równowagi ogólnej i autoregresja wektorowa. Aspekty praktyczne, *Przegląd Statystyczny,*(w druku).
[38] Wróbel-Rotter R., (2013c), Estymowane modele równowagi ogólnej i wektorowa autoregresja: model hybrydowy, *Bank i Kredyt*, (w druku).
[39] Wróbel-Rotter R., (2013d), Hybrydowy model wektorowej autoregresji – analiza empiryczna funkcji odpowiedzi na zakłócenia strukturalne, manuscript niepublikowany.
[40] Zellner A., (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Ltd., New York.

ESTYMOWANE MODELE RÓWNOWAGI OGÓLNEJ I AUTOREGRESJA WEKTOROWA. ASPEKTY TEORETYCZNE

Streszczenie

Model DSGE-VAR składa się z dwóch modeli autoregresji wektorowej: pierwszy z nich, pomocniczy, jest aproksymacją estymowanego modelu równowagi ogólnej, zapisanego w formie reprezentacji w przestrzeni stanów, i służy konstrukcji rozkładu *a priori* dla drugiego, szacowanego dla danych obserwowanych. Łączne wnioskowanie o parametrach modelu strukturalnego i autoregresyjnego jest możliwe po zbudowaniu odpowiednich rozkładów prawdopodobieństwa, stanowiących podstawę metod bayesowskich. Kluczową rolę pełni parametr wagowy, ustalający optymalne proporcje obydwu podejść i mający zasadnicze znaczenie dla oszacowania brzegowej gęstości obserwacji, stanowiącej podstawę do porównań mocy wyjaśniającej modeli. Artykuł stanowi syntezę informacji teoretycznych związanych z metodologią DSGE-VAR, i może być traktowany jako etap wstępny i wprowadzający w badania empiryczne.

Słowa kluczowe: DSGE-VAR, dynamiczny stochastyczny model równowagi ogólnej, wnioskowanie bayesowskie, specyfikacja rozkładu *a priori*

AN ESTIMATED GENERAL EQUILIBRIUM MODEL AND VECTOR AUTOREGRESSION. THEORETICAL ASPECTS

Abstract

The DSGE-VAR model consists of two models of vector autoregressions: the first one approximates linearised solution of the dynamic stochastic general equilibrium model and is used as a tool for construction of a prior distribution for the second one, estimated with the observed data. Combined inference is possible on the basis on probability distributions with the Bayesian techniques. The key role in the hybrid model is played by the weighting parameter that defines the relative proportions of the structural and autoregressive models. It has crucial impact for the marginal data density that allows to compare the power of different models. The main purpose of the paper is to present in details model assumptions and estimation.

Keywords: DSGE-VAR, dynamic stochastic general equilibrium model, Bayesian inference, prior specification

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

ALICJA OLEJNIK

WYBRANE METODY TESTOWANIA MODELI REGRESJI PRZESTRZENNEJ

1.WPROWADZENIE

Regresja liniowa jest jedną z najpopularniejszych technik eksploracji danych. Zakłada ona jednakże, iż obserwacje w próbie są niezależne, co niekoniecznie jest spełnione w przypadku danych przestrzennych, często charakteryzujących się autokorelacją przestrzenną. Współcześnie w literaturze światowej wiele miejsca poświęca się problemom analizy i modelowania zjawisk o charakterze przestrzennym, marginalizując jednak często aspekt weryfikacji poprawności założonej struktury autokorelacji przestrzennej. Powszechne podejście do modelowania zjawisk przestrzennych opiera się na technicznej eliminacji problemu autokorelacji przestrzennej z modelu ekonometrycznego. Nietrudno jednak zauważyć, że potraktowanie autokorelacji przestrzennej, jako dodatkowego źródła informacji, pomaga lepiej modelować rzeczywiste zjawiska o charakterze przestrzennym. Takie podejście wymaga jednak skutecznych narzędzi testujących i weryfikujących poprawność wyboru postaci modelu przestrzennego.

Celem niniejszego artykułu jest zatem przybliżenie wybranych metod pozwalających na wybór modeli najlepiej opisujących zjawiska o charakterze przestrzennym. W niniejszej pracy opisane zostaną podstawowe testy dla modeli regresji przestrzennej ze wskazaniem ich pozytywnych i negatywnych stron oraz zalecenia dotyczące ich stosowalności. Zaprezentowane metody pozwalają na testowanie występowania autokorelacji przestrzennej w modelu oraz wybór najlepszej specyfikacji modelu przestrzennego. W artykule wskazano również propozycję odpowiedniej procedury testującej przestrzenną niestacjonarność, mającej istotny wpływ na wybór ostatecznej postaci modelu.

2. NIESTANDARDOWY TEST MORANA *I* W MODELU REGRESJI LINIOWEJ Z AUTOKORELACJĄ PRZESTRZENNĄ SKŁADNIKA LOSOWEGO

Przypomnijmy, że autokorelacja przestrzenna danych oznacza stopień skorelowania obserwowanej wartości zmiennej z wartością tej samej zmiennej w innej lokalizacji w przestrzeni (np. regionie). Wyróżniamy dwa rodzaje autokorelacji przestrzennej: dodatnią i ujemną. Dla autokorelacji dodatniej obiekty, które są położone blisko siebie mają tendencję do przyjmowania podobnych wartości, dla ujemnej autokorelacji prze-

strzennej, obserwujemy, sąsiedztwo wysokich wartości z niskimi (i odwrotnie). Jeśli autokorelacja przestrzenna występuje w modelu ekonometrycznym, należy uwzględnić ją w postaci modelu, aby zminimalizować jej negatywny wpływ na jakość oszacowań parametrów modelu[1]. Zauważmy również, że przy dużej autokorelacji przestrzennej efektywna wielkość próby jest mniejsza niż obserwowana.

Rozważmy model z przestrzennie autoskorelowanym składnikiem losowym, który uwzględnia przestrzenną zależność w rozkładzie składnika losowego (por. Anselin, 1988). Wówczas poszczególne błędy u_i dla różnych lokalizacji są skorelowane z błędami w innych lokalizacjach:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{1}$$

gdzie **y** jest wektorem zmiennych losowych, z których każda reprezentuje odpowiedni region ekonomiczny (województwo, państwo, itp.), **W** oznacza przestrzenną macierz wag, o wymiarach $N \times N$, a przez macierz **X**, o wymiarach $N \times K$ rozumieć będziemy macierz zmiennych objaśniających, z uwzględnieniem wyrazu wolnego, wnoszącą do modelu dodatkową, zewnętrzną informację o badanym procesie. Zmienna **u** to N-wymiarowy wektor skorelowanych przestrzennie składników losowych, zaś **ε** to proces gaussowskiego białego szumu z zerową wartością oczekiwaną i macierzą wariancji-kowariancji $\sigma^2\mathbf{I}$. Parametr λ jest współczynnikiem przestrzennej korelacji reszt, dla którego przyjmujemy założenie: $|\lambda|<1$.

Test Morana opracowany został przez Patricka Morana w roku 1950 i jest najpowszechniej stosowanym testem autokorelacji przestrzennej. Poniżej przedstawimy test Morana jednak w nieco odmiennym od standardowego ujęciu. W odróżnieniu od podejścia klasycznego nie będziemy zakładać *a priori* normalności rozkładu składnika losowego, a raczej rozważymy asymptotyczny rozkład statystyk testowych. Kelejian i Prucha (2001) argumentują, że możemy przyjąć ogólne założenia, przy których statystyka Morana I jest asymptotycznie równoważna formie kwadratowej składnika losowego oraz podają ogólną postać Centralnego Twierdzenia Granicznego dla takiej formy kwadratowej. Te własności statystyki I pozwalają na uzależnienie elementów macierzy formy kwadratowej od wielkości próby, tak jak to zazwyczaj ma miejsce w przypadku modeli typu Cliffa-Orda (1981). Takie podejście, dopuszcza heteroskedastyczność składnika losowego. Należy podkreślić, iż wcześniejsze rozważania dotyczące asymptotycznego rozkładu statystyki I w teście Morana, lub równoważnych testów opartych na mnożniku Lagrange'a, nie uwzględniały przypadku, w którym elementy przestrzennej macierzy wag zależą od wielkości próby. Podobnie, nie dopuszczano heteroskedastyczności składnika losowego, przyjmując *a priori* jego rozkład normalny $\mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

[1] Przypomnijmy, że w przypadku występowania zjawiska autokorelacji przestrzennej estymator metody najmniejszych kwadratów może nie być zgodny, ani nawet nieobciążony (por. Anselin, 1988).

Poniżej zdefiniujemy statystykę testową *I* dla modelu regresji liniowej bez opóźnień przestrzennych w zmiennej zależnej. Następnie pokażemy, że test Morana oraz test mnożników Lagrange'a (*LM*) przestrzennej autokorelacji składnika losowego są tożsame (por. Burridge, 1980). Okaże się, iż statystyka testu Morana jest zdefiniowana w postaci formy kwadratowej składnika losowego (lub reszt) znormalizowanych przez szacowany błąd standardowy formy kwadratowej. Cliff i Ord (1973, 1981) w swojej definicji statystyki Morana *I* zastosowali czynnik normalizujący będący niezgodnym estymatorem odchylenia standardowego formy kwadratowej. Wymagał on dalszej normalizacji w celu uzyskania statystyki asymptotycznie standaryzowanej. W kolejnych publikacjach błąd ten był powtarzany przez wielu autorów.

2.1. DEFINICJA STATYSTYKI *MORANA I*

Niech **e** oznacza wektor reszt modelu z estymacji MNK. Wówczas, przy hipotezie zerowej H_0: $\lambda = 0$ o braku autokorelacji przestrzennej reszt modelu (z hipotezą alternatywną: H_1: $\lambda \neq 0$) statystyka Morana *I* przyjmuje postać:

$$I = \frac{\mathbf{e}^T \mathbf{W} \mathbf{e}}{N^{-1} \mathbf{e}^T \mathbf{e} [tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2}}, \tag{2}$$

gdzie *tr*(.) oznacza ślad macierzy.

Wyniki prac opublikowane przez Kelejiana i Pruchę (2001) dowodzą, iż przy założeniu prawdziwości hipotezy H_0 statystyka *I* dąży względem dystrybuanty do rozkładu normalnego, tj.:

$$I \xrightarrow{D} \mathrm{N}(0,1).$$

Na uwagę zasługuje fakt, iż w większości pozycji literaturowych, we wzorze (2) dzielnik, będący pierwiastkiem śladu macierzy zastępowany jest sumą elementów macierzy **W**. Ponieważ w ogólności:

$$[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2} \neq \Sigma_i \Sigma_j w_{ij}, \tag{3}$$

to powszechnie stosowane podejście jest niepoprawne. Należy nadmienić, że w swoim artykule z 1950 roku Moran rozważał przestrzenne macierze o wagach przyjmujących tylko wartości równe zero lub jeden. Wówczas zachodzi: $w_{ij} = w^2_{ij}$, co we wzorze (3) daje równość[2].

---

[2] Zauważmy, że własności testu takie jak np. moc, czy nieobciążoność zależą od stopnia w jakim macierz wag przestrzennych wykorzystana do wyliczenia statystyki Morana właściwie odwzorowuje rzeczywiste powiązania pomiędzy obiektami przestrzennymi. Problem ten dotyczy wszystkich testów autokorelacji przestrzennej składnika losowego i jest zauważany oraz aktualnie dyskutowany przez specjalistów (por. Drukker, Prucha, 2013).

2.2. TEST OPARTY NA MNOŻNIKACH LAGRANGE'A

Przy przyjętych powyżej oznaczeniach oraz założeniach łatwo zauważyć, że statystyka *LM* przyjmuje postać:

$$LM = \left[\frac{\mathbf{e}^T \mathbf{W} \mathbf{e}}{N^{-1} \mathbf{e}^T \mathbf{e} [tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2}} \right]^2 \sim \chi^2(1). \tag{4}$$

Zatem statystyka testowa *LM* jest kwadratem statystyki Morana *I*. Choć fakt ten został zaobserwowany już w 1980 roku przez Burridge'a (1980), do dziś pozostaje w cieniu szeroko rozpowszechnionego poglądu o rozłączności tychże dwóch testów.

2.3. STANDARYZACJA STATYSTYKI MORANA *I* DLA MAŁYCH PRÓB

Oznaczmy reszty modelu (1) przez: $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, gdzie $\hat{\boldsymbol{\beta}}$ jest estymatorem postaci: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Wprowadźmy oznaczenie: $\mathbf{M}_\mathbf{X} = \mathbf{I} - \mathbf{P}_\mathbf{X}$, gdzie $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Wówczas:

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\
&= [\mathbf{X}\boldsymbol{\beta} + \mathbf{u} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}] \\
&= \mathbf{u} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} \\
&= \mathbf{M}_\mathbf{X}\mathbf{u}.
\end{aligned}$$

Przy założeniu o normalności składnika losowego można pokazać, że:

$$E(I) = \frac{-N}{N-K} \frac{1}{[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2}} tr(\mathbf{P}_\mathbf{X}\mathbf{W}),$$

$$Var(I) = C\left[tr(\mathbf{M}_\mathbf{X}\mathbf{W}\mathbf{M}_\mathbf{X}\mathbf{W}^T) + tr(\mathbf{M}_\mathbf{X}\mathbf{W}\mathbf{M}_\mathbf{X}\mathbf{W}) + [tr(\mathbf{M}_\mathbf{X}\mathbf{W})]^2 \right] - E^2(I),$$

gdzie: $C = \frac{N^2}{(N-K)(N-K+2)} \frac{1}{[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]}$.

W przypadku małych prób, wartość oczekiwana statystyki Morana *I* może być zatem różna od zera, a jej wariancja różna od jedności. Dla małych prób można jednak zaproponować standaryzację, która umożliwi testowanie obecności autokorelacji przestrzennej.

Warto zauważyć, iż w literaturze spotyka się liczne uproszczenia postaci wariancji statystyki Morana *I*, przy przyjęciu pewnych dodatkowych warunków uszczegóławiających, np.: $\mathbf{W}^{\mathrm{T}}=\mathbf{W}$.

Przy pewnych założeniach dotyczących macierzy **W** zachodzi, iż:

$$\lim_{N\to\infty} E(I)=0, \quad \lim_{N\to\infty} Var(I)=1.$$

Ponieważ $I \xrightarrow{D} \mathrm{N}(0,1)$ możemy wnioskować, że:

$$\frac{I-E(I)}{[Var(I)]^{1/2}} \xrightarrow{D} \mathrm{N}(0,1).$$

I tak na przykład, dla testu Morana hipoteza zerowa o braku autokorelacji przestrzennej, przy poziomie istotności 0,05, odrzucana jest gdy:

$$\frac{I-E(I)}{[Var(I)]^{1/2}} \geq 1,96$$

3. NIESTANDARDOWY TEST MORANA *I* W MODELU AUTOREGRESJI PRZESTRZENNEJ Z AUTOKORELACJĄ PRZESTRZENNĄ SKŁADNIKA LOSOWEGO

Rozważmy teraz model autoregresji przestrzennej z autokorelacją przestrzenną składnika losowego:

$$\begin{aligned} &\mathbf{y}=\rho\mathbf{W}\mathbf{y}+\mathbf{X}\boldsymbol{\beta}+\mathbf{u}, \\ &\mathbf{y}=\mathbf{Z}\boldsymbol{\delta}+\mathbf{u}, \\ &\mathbf{u}=\lambda\mathbf{W}\mathbf{u}+\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}\sim \mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}), \end{aligned} \tag{5}$$

gdzie $|\rho|<1$ – współczynnikiem przestrzennym, $|\lambda|<1$ – współczynnikiem przestrzennej autokorelacji składnika losowego oraz $\mathbf{Z}=[\mathbf{X},\mathbf{W}\mathbf{y}]$, $\boldsymbol{\delta}=[\boldsymbol{\beta}^T,\rho]^T$.

Przy standardowych założeniach możemy rozważać estymator Metody Zmiennych Instrumentalnych dla parametru $\boldsymbol{\delta}$:

$$\hat{\boldsymbol{\delta}}=(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}^T\mathbf{y},$$

gdzie: $\hat{\mathbf{Z}} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Z}$, $\mathbf{H}$ – macierz instrumentów, oraz $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\delta}}$, oznacza wektor reszt modelu. Stawiamy hipotezę zerową H_0: $\lambda \neq 0$, przy hipotezie alternatywnej H_1: $\lambda \neq 0$ Statystyka Morana *I* dla modelu autoregresji przestrzennej z autokorelacją przestrzenną składnika losowego przyjmuje postać:

$$I = \frac{\mathbf{e}^T\mathbf{W}\mathbf{e}}{\left[(N^{-1}\mathbf{e}^T\mathbf{e})^2\,[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})] + (N^{-1}\mathbf{e}^T\mathbf{e})\hat{b}^T\hat{b}\right]^{1/2}}, \tag{6}$$

gdzie: $\hat{b} = -\mathbf{H}(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1}\mathbf{Z}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{Z}^T(\mathbf{W} + \mathbf{W}^T)\mathbf{e}$, oraz $I \xrightarrow{D} \mathrm{N}(0,1)$[3].

Zatem przy dwustronnym teście na poziomie istotności 5% hipoteza zerowa H_0: $\lambda = 0$ jest odrzucana dla wartości $|I| > 1{,}96$.

4. TEST F DLA MODELU PRZESTRZENNEGO

Analogicznie jak dla modeli klasycznych, do porównywania oszacowanych przestrzennych modeli ekonometrycznych można wykorzystać test *F*. Rozważmy zatem dwa przykładowe modele przestrzenne, przestrzenny model autoregresyjny SAR oraz przestrzenny model SADL (Spatially autoregressive-distributed lag model) (Olejnik, 2008):

$$\text{SAR}: \quad \mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$

$$\text{SADL}: \quad \mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$

gdzie $\mathbf{X}$ to macierz zmiennych objaśniających o wymiarach ($N \times K$), $\boldsymbol{\beta}$ to odpowiadający jej *k*-elementowy wektor parametrów, natomiast $\boldsymbol{\gamma}$ jest *k*-elementowym wektorem parametrów dla ważonej przestrzennie macierzy zmiennych objaśniających. Przyjmijmy założenie, że $\boldsymbol{\beta} \neq \mathbf{0}$. W celu zbadania istotności wektora przestrzennego $\boldsymbol{\gamma}$ stawiamy hipotezę: H_0: $\gamma = \mathbf{0}$ przy hipotezie alternatywnej H_1: $\gamma \neq \mathbf{0}$. Statystyka *F* przyjmuje wówczas postać:

$$F = \frac{N - (2K+1)}{K+1}\,\frac{\left[RSS_{\text{SAR}} - RSS_{\text{SADL}}\right]}{RSS_{\text{SADL}}} \sim F(K+1, N-(2K+1)),$$

[3] Rozumowanie pozwalające stwierdzić, że statystyka o postaci (6) dąży według dystrybuanty do rozkładu normalnego można znaleźć w opracowaniu Kelejian i Prucha (2001).

gdzie RSS_{SADL} oznacza sumę kwadratów reszt z modelu SADL, zaś RSS_{SAR} – sumę kwadratów reszt modelu SAR.

Za pomocą testu *F* istnieje również możliwość wykonania analizy porównawczej dla dowolnych dwóch modeli przestrzennych. Zauważmy, że analogiczne rozumowanie możemy przeprowadzić również dla modelu przestrzennego i klasycznej regresji liniowej. Taka analiza w istotny sposób może pomóc w podjęciu decyzji o zasadności, bądź nie zastosowania modelu przestrzennego. W ocenie autora test ten wydaje się być niedoceniany przez badaczy pomimo swej prostej konstrukcji i istotnej roli dla wyboru właściwej postaci modelu. Za pomocą powyższego testu można dokonywać porównań nie tylko pomiędzy modelami o takiej samej postaci funkcyjnej, ale również różniących się macierzami wag przestrzennych **W**.

5. TEST J DLA SPECYFIKACJI MODELU PRZESTRZENNEGO

Poniżej zaprezentujemy test oparty na procedurze testu *J*, będącej znanym algorytmem testowania modeli niezagnieżdżonych (por. Kelejian, 2008). Przedstawiony tu test, umożliwia porównanie wybranego modelu przestrzennego z pojedynczym lub wieloma niezagnieżdżonymi modelami alternatywnymi. Rozważane modele mogą, choć nie muszą zawierać opóźnienia przestrzenne, zarówno w zmiennej zależnej, jak i składniku losowym. Idea testu sprowadza się do odpowiedzi na pytanie, czy model alternatywny wnosi dodatkową informację w stosunku do modelu wyjściowego.

Rozważmy autoregresyjny model przestrzenny z autokorelacją składnika losowego. Jako hipotezę zerową, zaproponujmy model wyjściowy:

$$H_0: \left| \begin{array}{l} \mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \quad = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \\ \mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{array} \right| . \tag{8}$$

Jako alternatywę, przyjmujemy *G* przeciwstawnych modeli:

$$H_i: \left| \begin{array}{l} \mathbf{y} = \rho_i \mathbf{W}_i \mathbf{y} + \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{u}_i \\ \quad = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{u}_i \\ \mathbf{u}_i = \lambda_i \mathbf{W}_i \mathbf{u}_i + \boldsymbol{\varepsilon}_i \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{array} \right| , \tag{9}$$

gdzie $i = 1, ..., G$. Zauważmy, że w szczególnym przypadku, model wyjściowy od modeli alternatywnych może różnić się tylko postacią przestrzennej macierzy wag, co z punktu widzenia analiz przestrzennych jest kwestią bardzo istotną.

W wyniku przemnożenia obustronnego przez macierz $(\mathbf{I}-\lambda\mathbf{W})$ modelu (8) możemy go zapisać w postaci:

$$(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}=(\mathbf{I}-\lambda\mathbf{W})\mathbf{Z}\boldsymbol{\delta}+\boldsymbol{\varepsilon}. \tag{10}$$

Dla skrócenia zapisu, przyjmijmy następujące oznaczenia: $\mathbf{y}(\lambda) = (\mathbf{I}-\lambda\mathbf{W})\mathbf{y}$, $\mathbf{Z}(\lambda) = (\mathbf{I}-\lambda\mathbf{W})\mathbf{Z}$, $\mathbf{Z}_i(\lambda_i) = (\mathbf{I}-\lambda_i\mathbf{W}_i)\mathbf{Z}_i$. Wówczas model (10) przyjmie postać:

$$\mathbf{y}(\lambda)=\mathbf{Z}(\lambda)\boldsymbol{\delta}+\boldsymbol{\varepsilon}. \tag{11}$$

Przez $\hat{\boldsymbol{\delta}}_i$ oznaczmy pewien estymator parametru $\boldsymbol{\delta}_i$. Wówczas modyfikacja klasycznego testu *J* (por. Davidson i MacKinnon, 1981; Pesaran i Weeks, 2001) na przypadek przestrzenny wymaga rozszerzenia modelu (11) do postaci:

$$\begin{aligned}\mathbf{y}(\lambda)&=\mathbf{Z}(\lambda)\boldsymbol{\delta}+\sum_{i=1}^{G}\alpha_i[\mathbf{Z}_i(\lambda_i)\hat{\boldsymbol{\delta}}_i]+\boldsymbol{\varepsilon}\\&=\mathbf{Z}(\lambda)\boldsymbol{\delta}+\sum_{i=1}^{G}\alpha_i[\mathbf{Z}_i\hat{\boldsymbol{\delta}}_i]+\sum_{i=1}^{G}\phi_i[\mathbf{W}_i\mathbf{Z}_i\hat{\boldsymbol{\delta}}_i]+\boldsymbol{\varepsilon}.\end{aligned} \tag{12}$$

gdzie α_i, ϕ_i, są parametrami oraz spełniony jest warunek: $\phi_i=-\alpha_i\lambda_i$. Zauważmy, że dla modelu (8), z hipotezy zerowej mamy: $\alpha_i = 0$. Na mocy powyższego możemy przeformułować testowaną hipotezę:

$$H_0\colon \alpha_i=\phi_i=0, \text{ dla } i=1,\ldots,G \tag{13}$$

W swojej pracy Kelejian (2008), jako procedurę estymacyjną, zaproponował metodę opartą na zastosowaniu zmiennych instrumentalnych. Rozważmy zatem następujące macierze instrumentów:

$$\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}, \ldots, \mathbf{W}^r\mathbf{X}]_{LNZ},$$

$$\mathbf{H}_i = [\mathbf{X}_i, \mathbf{W}_i\mathbf{X}_i, \mathbf{W}_i^2\mathbf{X}_i, \ldots, \mathbf{W}_i^r\mathbf{X}_i]_{LNZ},$$

$$\overline{\mathbf{H}}_i = [\overline{\mathbf{X}}, \mathbf{W}\overline{\mathbf{X}}, \mathbf{W}^2\overline{\mathbf{X}}, \ldots, \mathbf{W}^r\overline{\mathbf{X}}]_{LNZ},$$

gdzie $\overline{\mathbf{X}} = [\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_G]$, natomiast indeks *LNZ*, oznacza liniowo niezależne kolumny rozważanej macierzy[4]. Zakłada się, iż przyjęte macierze instrumentów zawierają co najmniej wszystkie liniowo niezależne kolumny (w sensie rozpina-

[4] W większości przypadków nie rozważa się instrumentów z macierzą wag w potędze wyższej niż dwa.

nej przestrzeni) odpowiednich macierzy: $[\mathbf{X}, \mathbf{WX}]$, $[\mathbf{X}_i, \mathbf{W}_i\mathbf{X}_i]$, oraz $[\overline{\mathbf{X}}, \mathbf{W}\overline{\mathbf{X}}]$, dla $i = 1, \ldots, G$. Przy powyższych oznaczeniach, możemy przedstawić etapy procedury postępowania w przypadku testu *J* dla modeli przestrzennych.

Etap 1

Estymujemy model (8) metodą 2MNK (dwustopniową MNK) przy użyciu macierzy instrumentów **H**, skąd otrzymujemy reszty modelu, który oznaczmy przez $\hat{\mathbf{u}}$. Równolegle, estymujemy *G* modeli alternatywnych postaci (9), metodą 2MNK, przy użyciu macierzy instrumentów $\mathbf{H}_i$, uzyskując oszacowania parametrów $\boldsymbol{\delta}_i$.

Etap 2

Stosujemy procedurę UMM (Uogólnionej Metody Momentów) dla modelu (8), w celu estymacji parametru λ z wykorzystaniem wyliczonych wcześniej reszt modelu $\hat{\mathbf{u}}$. Podstawiając oszacowany parametr $\hat{\lambda}$ do modelu (11) należy oszacować tak otrzymany model metodą 2MNK przy zastosowaniu macierzy instrumentów **H**. Następnie, otrzymany z estymacji wektor reszt $\hat{\boldsymbol{\varepsilon}}$, wykorzystujemy do estymacji wariancji składnika losowego:

$$\hat{\sigma}_{\varepsilon}^2 = N^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}. \tag{14}$$

Etap 3

W rozszerzonym modelu (12) następujemy parametr λ jego oszacowaniem $\hat{\lambda}$, otrzymanym w poprzednim kroku.

Przyjmijmy oznaczenia:

$$\Upsilon = [\mathbf{Z}_1\hat{\boldsymbol{\delta}}_1, \ldots, \mathbf{Z}_G\hat{\boldsymbol{\delta}}_G, \mathbf{W}_1\mathbf{Z}_1\hat{\boldsymbol{\delta}}_1, \ldots, \mathbf{W}_G\mathbf{Z}_G\hat{\boldsymbol{\delta}}_G], \psi = [\alpha_1, \ldots, \alpha_G, \phi_1, \ldots, \phi_G]^T.$$

Wówczas empirycznym odpowiednikiem rozszerzonego modelu (12) będzie model postaci:

$$\mathbf{y}(\hat{\lambda}) \approx \mathbf{Z}(\hat{\lambda})\boldsymbol{\delta} + \Upsilon\boldsymbol{\psi} + \boldsymbol{\varepsilon}. \tag{15}$$

Etap 4

W kolejnym kroku należy oszacować parametry modelu (15) metodą 2MNK z zastosowaniem macierzy instrumentów $\overline{\mathbf{H}}$. Oznaczmy macierz zmiennych objaśniających w modelu (15) przez $\Xi \equiv [\mathbf{Z}(\hat{\lambda}), \Upsilon]$, parametry regresji zaś przez $\boldsymbol{\gamma}^T \equiv [\boldsymbol{\delta}^T, \boldsymbol{\psi}^T]$. Zauważmy, że dla modelu określonego w hipotezie zerowej mamy $\boldsymbol{\gamma}_0^T = [\boldsymbol{\delta}^T, \mathbf{0}^T]$. Niech $\hat{\Xi} = \mathbf{P}_{\overline{\mathbf{H}}}\Xi \equiv [\hat{\mathbf{Z}}(\hat{\lambda}), \hat{\Upsilon}]$, gdzie operator rzutu na macierz Ξ jest postaci: $\mathbf{P}_{\overline{\mathbf{H}}} = \overline{\mathbf{H}}(\overline{\mathbf{H}}^T\overline{\mathbf{H}})^{-1}\overline{\mathbf{H}}^T$ Wówczas estymator 2MNK parametru γ przyjmuje postać:

$$\hat{\boldsymbol{\gamma}} = (\hat{\Xi}^T\hat{\Xi})^{-1}\hat{\Xi}^T\mathbf{y}(\hat{\lambda}). \tag{16}$$

Twierdzenie

Przy pewnych standardowych założeniach (por. Kelejian, 2008) i przyjętych powyżej oznaczeniach, mamy:

$$\mathrm{N}^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\xrightarrow{D}\mathrm{N}(\mathbf{0},\hat{\sigma}_{\varepsilon}^2\operatorname*{plim}_{N\to\infty}(\hat{\Xi}^T\hat{\Xi})^{-1}),$$

$$\sigma_{\varepsilon}^2=\operatorname*{plim}_{N\to\infty}\hat{\sigma}_{\varepsilon}^2,$$

gdzie plim oznacza zbieżność względem prawdopodobieństwa[5].

Na mocy powyższego twierdzenia, wnioskowania dla małych prób mogą być oparte na założeniu:

$$\hat{\boldsymbol{\gamma}}\ \simeq\ \mathrm{N}\left(\boldsymbol{\gamma},\hat{\sigma}_{\varepsilon}^2(\hat{\Xi}^T\hat{\Xi})^{-1}\right). \tag{17}$$

Niech $\overline{K}\equiv K+1+2G$, $\hat{\boldsymbol{\gamma}}^T\equiv[\hat{\boldsymbol{\delta}}^T,\hat{\boldsymbol{\psi}}^T]$ oraz niech $\hat{\mathbf{VC}}_{\hat{\psi}}$ reprezentuje oszacowanie macierzy wariancji-kowariancji estymatora $\hat{\boldsymbol{\psi}}$ opartego na małej próbie. Wówczas $\hat{\mathbf{VC}}_{\hat{\psi}}$ stanowi podmacierz o wymiarach 2G×2G macierzy wariancji-kowariancji parametru $\hat{\boldsymbol{\gamma}}$ – macierzy $\hat{\sigma}_{\varepsilon}^2(\hat{\Xi}^T\hat{\Xi})^{-1}$ o wymiarach $\overline{K}\times\overline{K}$. Wówczas test Walda (por. Godfrey i Pesaran, 1983) odrzuca hipotezę zerową H_0: $\boldsymbol{\psi}=\mathbf{0}$ na korzyść hipotezy alternatywnej H_1: $\boldsymbol{\psi}\neq\mathbf{0}$ na poziomie istotności α, gdy:

$$\left(\hat{\boldsymbol{\psi}}^T\hat{\mathbf{VC}}_{\hat{\psi}}^{-1}\hat{\boldsymbol{\psi}}\right)>\chi_{1-\alpha}^2(2G), \tag{18}$$

gdzie $\chi^2_{(1-\alpha)}(2G)$ oznacza kwantyl rozkładu χ^2 o 2G stopniach swobody, rzędu 1-α.

6. TESTOWANIE STACJONARNOŚCI PRZESTRZENNEJ

Problem niestacjonarności przestrzennej jako pierwszy rozważał Fingleton (1999). W swej pracy zauważył on, że przestrzenna niestacjonarność, analogicznie jak w przypadku szeregów czasowych, może prowadzić do wystąpienia regresji pozornej. Poniżej zaprezentowano procedurę testową umożliwiającą wykluczenie występowania problemu niestacjonarności przestrzennej.

Jako punkt wyjścia przyjmijmy model postaci:

$$\mathbf{y}=\boldsymbol{\rho}\mathbf{W}\mathbf{y}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim\mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}). \tag{19}$$

[5] Dowód powyższego twierdzenia można znaleźć w pracy Kelejian (2008).

W powyższym równaniu spełnienie warunku $|\rho|<1$ zapewnia co najmniej asymptotyczną stacjonarność danych generujących proces przestrzenny. W tym wypadku możemy powiedzieć, że **y** jest przestrzennie zintegrowanym procesem rzędu zero – SI(0).

Rozważmy model postaci:

$$\mathbf{y}=\mathbf{X\beta}+\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}=\lambda\mathbf{W\varepsilon}+\boldsymbol{\mu}, \quad \boldsymbol{\mu}\sim \mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}). \tag{20}$$

Kosfeld i Lauridsen (2004) zaproponowali dwustopniowy test wskazujący na istnienie przestrzennej stacjonarności. Ich strategia, oparta jest na dwustopniowym teście mnożników Lagrange'a dla przestrzennie skorelowanego składnika losowego[6]. Zatem dla modelu , przy H_0: $\lambda = 0$ możemy zastosować procedurę testu *LM* dla reszt modelu (LME). Jeśli statystyka testowa okaże się nieistotnie różna od zera, powiemy, że nie mamy tu do czynienia z autokorelacją przestrzenną ($\lambda=0$). W przeciwnym wypadku możemy wnioskować o występowaniu autokorelacji przestrzennej lub też o istnieniu przestrzennej niestacjonarności.

W drugim etapie, należy zweryfikować, czy parametr przestrzennej autoregresji w równaniu (20) wynosi jeden, czy jest mniejszy od jedności. W tym celu należy zastosować test *LM* dla przestrzennie zróżnicowanego modelu (DLME) postaci:

$$\Delta\mathbf{y}=\Delta\mathbf{X\beta}+\boldsymbol{\mu} \quad \boldsymbol{\mu}\sim \mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}), \tag{21}$$

gdzie: $\mu=(1-\lambda\mathbf{W})\boldsymbol{\varepsilon}$. Wówczas hipoteza zerowa jest równoważna: H_0:$|\lambda|=1$, przy hipotezie alternatywnej H_1:$|\lambda|<1$.

Analogiczna procedura może zostać zastosowana w celu badania przestrzennej niestacjonarności każdej ze zmiennych rozważanego modelu. Wtedy należy wykonać regresję danej zmiennej na czynnik stały. Bardziej szczegółową analizę problemu przestrzennej niestacjonarności wraz z przykładem badania przestrzennej stacjonarności dla procesu konwergencji w UE można znaleźć w publikacji autora (Olejnik, 2008).

---

[6] Własności testu zostały dokładnie opisane w pracy Kosfelda i Lauridsena (2006). Alternatywne metody testowania można znaleźć w pracy Kosfelda i Lauridsena (2004), gdzie zastosowana procedura oparta została na teście Walda oraz w opracowaniu Lauridsena (1999), w którym wykorzystano podejście Dickey-Fullera.

7. PODSUMOWANIE

W artykule przedstawiono metody testowania autokorelacji przestrzennej w najpopularniejszych modelach przestrzennych. Jako narzędzie do porównywania dwóch ekonometrycznych modeli przestrzennych, lub modelu przestrzennego z klasycznym zaproponowano zastosowanie znanego testu *F*. Z bardziej zaawansowanych narzędzi wspomagających dokonanie wyboru właściwego modelu zaprezentowano test *J*. Mimo dość skomplikowanej konstrukcji testu, jego rola dla oceny poprawności przyjętej struktury przestrzennej jest bardzo duża. Zdaniem autora, w chwili obecnej stanowi on najlepsze narzędzie porównawcze nieposiadające porównywalnej alternatywy. Ostatnia część pracy przybliża problem przestrzennej niestacjonarności oraz prezentuje odpowiednią procedurę testującą.

Mimo szerokich możliwości przedstawionych metod można zauważyć konieczność dalszego rozszerzania i modyfikacji zaprezentowanej metodologii. W dalszych opracowaniach planowane jest przedstawienie zatem konstrukcji testów dla przestrzennych modeli wielowymiarowych. W szczególności planuje się opracowanie wielowymiarowego testu *J* dla modeli klasy WAMP – Wielowymiarowych autoregresyjnych modeli przestrzennych (por. Olejnik, 2013).

Uniwersytet Łódzki

LITERATURA

[1] Anselin L., (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publications, Dordrecht.
[2] Burridge P., (1980), On the Cliff–Ord Test for Spatial Correlation, *Journal of the Royal Statistical Society*, B, 42 (1), 107–108.
[3] Cliff A., Ord J. K., (1981), *Spatial Processes: Models and Applications*, Pion, London.
[4] Drukker D. M., Prucha I. R., (2013), Finite Sample Properties of the $I^2(q)$ Test Statistic for Spatial Dependence, *Spatial Economic Analysis*, 8, 271–292.
[5] Fingleton B., (1999), Spurious Spatial Regression: Some Monte Carlo Results with Spatial Unit Root and Spatial Cointegration, *Journal of Regional Science*, 39, 1–19.
[6] Godfrey L., Pesaran H., (1983), Tests of Nonnested Regression Models After Estimation by Instrumental Variables or Least Squares, *Journal of Econometrics*, 21, 133–154.
[7] Kelejian H. H., (2008), A Spatial Test J for Model Specification Against a Single or a Set of Non–Nested Alternatives, *Letters in Spatial and Resource Sciences*, 1 (1), 3–11.
[8] Kelejian H. H., Prucha I. R., (2001), On the Asymptotic Distribution of the Moran I Test Statistic with Applications, *Journal of Econometrics*, 104, 219–257.
[9] Kosfeld R., Lauridsen J., (2004), Dynamic Spatial Modeling of Regional Convergence Processes, *Empirical Economics*, 29, 705–722.
[10] Kosfeld R., Lauridsen J., (2006), A Test Strategy for Spurious Regression, Spatial Nonstationarity, and Spatial Cointegration, *Papers in Regional Science*, 85 (3), 363–377.
[11] Lauridsen J., (1999), Spatial Cointegration Analysis in Econometric Modelling, *ERSA conference papers ersa99pa181*, European Regional Science Association.
[12] Moran P., (1950), Notes on Continuous Stochastic Phenomena, *Biometrika*, 37, 17–23.

[13] Olejnik A., (2008), Using the Spatial Autoregressively Distributed Lag Model in Assessing the Regional Convergence of per–capita Income in the EU25, *Papers in Regional Science*, 87 (3), 371–384.
[14] Olejnik A., (2013), Spatial Autoregressive Model – a Multidimensional Perspective with an Example Study of the Spatial Income Process in the EU 25, PREPARE, working papers, wysłane do recenzji do *Geographical Analysis*.
[15] Pesaran H., Weeks M., (2001), Nonnested Hypothesis Testing: An Overview, w: Baltagi B., (red.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

WYBRANE METODY TESTOWANIA MODELI REGRESJI PRZESTRZENNEJ

Streszczenie

Celem niniejszej pracy jest przybliżenie nowoczesnych metod testowania modeli regresji przestrzennej pozwalających wybrać poprawną postać struktury zależności przestrzennych. Metody te umożliwiają testowanie występowania autokorelacji przestrzennej, a w dalszym etapie pozwalają na właściwą specyfikację modelu przestrzennego. W artykule wskazano wady i zalety prezentowanych metod oraz przedstawiono pewne rekomendacje dotyczące ich stosowania. W opracowaniu przedstawiono również problem przestrzennej niestacjonarności oraz zaproponowano właściwą procedurę weryfikacyjną.

Słowa kluczowe: modele regresji przestrzennej, test *J*, test *F*, Statystyka Morana *I*, przestrzenna niestacjonarność

SOME TESTING METHODS FOR SPATIAL REGRESSION MODELS

Abstract

The aim of this paper is to introduce modern testing methods for Spatial Regression Models used for selection of the correct structure of spatial dependencies. Presented methodology allows one to test for the presence of spatial autocorrelation and at a further stage to make the correct model specification. The paper presents some merits and drawbacks of selected statistical test widely used in spatial modelling with some recommendations. The problem of spatial non-stationarity with an appropriate testing procedure is also introduced.

Keywords: spatial regression models, *J* test, *F* test, Moran *I*, spatial non-stationarity

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

GRZEGORZ SZAFRAŃSKI

SYNTETYCZNE WSKAŹNIKI WYPRZEDZAJĄCE W PROGNOZOWANIU INFLACJI W POLSCE[1,2]

1. WSTĘP

Prognozowanie inflacji jest jednym z podstawowych zadań banków centralnych, szczególnie tych, które realizując strategię bezpośredniego celu inflacyjnego deklarują utrzymywanie tempa zmian cen konsumenta w określonym paśmie wahań[3]. Takie zainteresowanie władz monetarnych jest konieczne w sytuacji, gdy nominalna i realna sfera gospodarki reagują z opóźnieniem na działanie instrumentów polityki pieniężnej. Tymczasem na inflację oddziałuje bardzo wiele czynników i celem tego badania jest sprawdzenie, które zmienne lub grupy zmiennych niosą z odpowiednim wyprzedzeniem użyteczną informację o zmianach cen koszyka dóbr konsumenta. Zmienne te nazywane są wskaźnikami wyprzedzającymi lub wiodącymi (ang. *leading indicators*).

W artykule sprawdzono, jakie korzyści daje zastosowanie metody wskaźników wyprzedzających do systematycznego, comiesięcznego prognozowania podstawowego wskaźnika inflacji w Polsce tj. indeksu cen towarów i usług konsumpcyjnych (dalej CPI). Postawiono roboczą hipotezę, że za pomocą metod statystycznych (analiza korelacji) można wskazać zmienne, których wahania poprzedzają zmiany inflacji w Polsce. W eksperymencie prognostycznym obejmującym okres od grudnia 2010 r. do listopada 2012 r. zastosowano proste (liniowe i jednorównaniowe) modele regresji, dalej zwane modelami wskaźników wyprzedzających[4]. Do prognoz wskaźnika inflacji w horyzoncie od 1 do 12 miesięcy naprzód wykorzystano jedynie informacje

---

[1] Artykuł jest znacznie zmienioną wersją artykułu Szafrański (2011) opublikowanego w roboczej wersji w Materiałach i Studiach NBP. Zmiany dotyczą przede wszystkim części empirycznej tj. celu badania, zastosowanych metod i zakresu próby, a także motywacji i przeglądu literatury. Artykuł nie wyraża ani poglądów ani oficjalnego stanowiska Narodowego Banku Polskiego.

[2] Podziękowania dla współpracowników NBP za okazaną pomoc, a także dla uczestników seminariów naukowych (IE NBP i Katedry Ekonometrii UŁ) za motywującą dyskusję i dla anonimowego recenzenta za cenne uwagi do ostatecznej wersji tekstu.

[3] W Polsce bank centralny stosuje strategię bezpośredniego celu inflacyjnego od 1999 roku.

[4] W artykule Szafrański (2011) sprawdzono, że modele te dostarczają na ogół równie dokładnych lub dokładniejszych punktowych prognoz CPI niż takie modele wzorcowe jak np. model oparty o autoregresję.

dostępne w chwili sporządzania prognoz (tzw. dane w czasie rzeczywistym). Aby ocenić korzyści, jakie może przynieść redukcja informacji pochodzącej z dużego zbioru danych, jakość prognoz na podstawie wskaźników indywidualnych porównano z prognozami opartymi na wskaźnikach syntetycznych. Wskaźniki syntetyczne zostały określone jako czynniki opisujące wspólną część zmienności grupy zmiennych i uzyskane metodą analizy głównych składkowych. Elementem dodanym tego artykułu, w porównaniu do innych badań dotyczących prognozowania inflacji w Polsce na podstawie dużych zbiorów danych, jest sprawdzenie, jaki wpływ na jakość uzyskiwanych prognoz ma wstępny dobór zmiennych do analizy czynnikowej na podstawie kryteriów statystycznych. Celem pośrednim badania jest również ustalenie zawartości informacyjnej wybranych wskaźników z punktu widzenia przewidywania inflacji. Oczywiście ustalenie tego typu relacji ma charakter głównie symptomatyczny, a nie przyczynowo-skutkowy. Jednak poszukiwanie natury tej relacji może mieć duże znaczenie dla określenia źródeł i genezy zmian cen w myśl idei, że jednym z kryteriów potwierdzenia związku przyczynowo-skutkowego pomiędzy zmiennymi jest możliwość uzyskania na jego podstawie dokładnej prognozy.

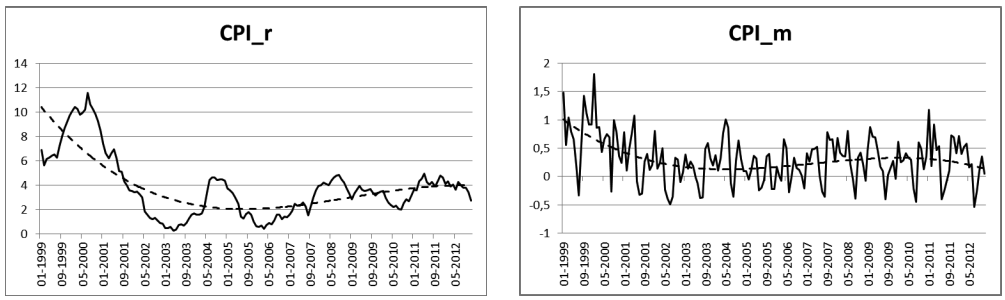
Struktura opracowania jest następująca. W drugiej części przedstawiono fakty dotyczące przebiegu inflacji w Polsce po okresie transformacji i postawiono szczegółowe hipotezy badawcze. Następnie wyjaśniono genezę koncepcji wskaźników wyprzedzających i skomentowano jej wykorzystanie w dotychczasowych badaniach nad prognozowaniem inflacji w Polsce i na świecie. W części trzeciej opisano konstrukcję modeli wskaźników wyprzedzających, a także szczegółowe założenia eksperymentu prognostycznego. Część czwarta zawiera opis konstrukcji indywidualnych i syntetycznych wskaźników wyprzedzających. W części piątej oceniono korzyści wynikające z ich stosowania oraz podjęto próbę oceny ich zawartości ekonomicznej.

2. INFLACJA W POLSCE PO OKRESIE TRANSFORMACJI – MOTYWACJA BADAWCZA

W latach 90. XX w. dwucyfrowa inflacja (w ujęciu rocznym) była w Polsce zjawiskiem powszechnym. Proces stopniowego jej spadku (dezinflacji) postępował wraz z wprowadzaniem nowych, systemowych rozwiązań w polityce pieniężnej. Trwały spadek inflacji poniżej 10% w skali roku i jej ustabilizowanie na średnim poziomie ok. 3% odnotowano dopiero po 2001 roku, a więc w dwa lata po wprowadzeniu strategii bezpośredniego celu inflacyjnego (por. rysunek 1).

Jednak po 2001 roku, wbrew temu, co pokazują statystyki opisowe wskaźnika CPI w ujęciu rok do roku (dalej CPI r/r) zmienność inflacji wcale nie uległa zmniejszeniu (por. tabela 1). Wskaźnik zmian cen w ujęciu miesięcznym (CPI m/m) ma bowiem zbliżoną wariancję w dwóch podpróbach obejmujących lata 2001–2006 i 2007–2012. Ponadto porównując wskaźniki publikowane ze wskaźnikami odsezonowanymi za pomocą procedury Tramo/Seats można spostrzec, że wskaźnik CPI m/m jest w dużym stopniu pod wpływem wahań sezonowych, które w latach 2001–2012 stanowiły

prawie 55% wariancji tego wskaźnika[5]. Zatem po wyłączeniu w miarę regularnych wahań sezonowych zmienność, która pozostaje do objaśnienia za pomocą modeli ekonometrycznych, mierzona odchyleniem standardowym, jest rzędu 0,25 pp. wobec ok. 0,36 pp. całkowitej zmienności inflacji w skali miesiąca (por. wyniki w tabeli 1). Wskazuje to, jak ważne, gdy badamy zmienność cen, jest poprawne uwzględnienie zmian o charakterze sezonowym.



Rysunek 1. Wskaźniki inflacji (CPI) w ujęciu rocznym (lewy panel) i miesięcznym (prawy panel) za okres 1999–2012 (w pp.) na tle nieliniowego trendu (linia przerywana)
Źródło: GUS, obliczenia własne.

Tabela 1.

Statystyki opisowe wskaźników inflacji (CPI) w ujęciu rocznym (w pp.), miesięcznym i po odsezonowaniu w latach 2001–2012 i w dwóch podokresach

Okres	statystyki opisowe	Indeks CPI r/r	Indeks CPI m/m	Odsezonowany CPI m/m	Wskaźnik sezonowości
2001–2012	Średnia	2,967	0,230	0,230	-0,000
	Odchylenie standardowe	1,540	0,363	0,246	0,267
2001–2006	Średnia	2,479	0,171	0,173	-0,001
	Odchylenie standardowe	1,884	0,358	0,264	0,268
2007–2012	Średnia	3,468	0,289	0,221	0,000
	Odchylenie standardowe	0,864	0,361	0,211	0,267

Wskaźniki sezonowości uzyskane w pakiecie Demetra+ 4.3 przy użyciu Tramo/Seats.
Źródło: obliczenia własne.

[5] Przy założeniu, że wskaźnik sezonowości nie jest skorelowany ze wskaźnikiem CPI wyrównanym sezonowo.

Pomimo, że zmienność inflacji wyrównanej sezonowo jest stosunkowo mała, to jej źródła nie są jednorodne. Potwierdzeniem tego może być porównanie częstości występowania nietypowych obserwacji w CPI ogółem i jego komponentach wykrytych w automatycznej procedurze Tramo. O ile w przebiegu szeregu CPI ogółem obserwacje nietypowe zdarzają się stosunkowo rzadko, to w jego komponentach występują one o wiele częściej[6]. Pomimo, że wpływ odstających obserwacji w komponentach jest wygładzany w indeksie CPI ogółem za pomocą wag, to nadal te nietypowe obserwacje mogą mieć w pojedynczych okresach istotny wpływ na kształtowanie się całego agregatu. Dodatkowo skład koszyka dóbr i usług nabywanych przez gospodarstwa domowe podlega co roku zmianom ocenianym na podstawie badań budżetów gospodarstw domowych. Wszystkie wymienione tu elementy dotyczące pomiaru inflacji w Polsce składają się na obraz podstawowych problemów, jakie można napotkać w systematycznym prognozowaniu wskaźnika CPI.

Na podstawie tego krótkiego przeglądu zmienności wskaźnika cen konsumpcyjnych po 2001 roku postawiono następującą hipotezę. Ze względu na wiele potencjalnych źródeł zmienności cen koszyka konsumenta uwzględnienie dużego zbioru informacji jest konieczne, żeby móc systematycznie prognozować inflację w sposób wiarygodny i trafny. Jednak w klasycznej analizie regresji na podstawie liniowych modeli (np. modelu indywidualnych wskaźników wiodących) oczekiwany błąd prognozy jest proporcjonalny do ilorazu liczby zmiennych objaśniających (predyktorów) i liczby obserwacji (N/T) – por. Stock i Watson (2006). Gdy liczba wybranych predyktorów przekracza kilkanaście zmiennych i jest niewiele większa od liczby obserwacji, błąd estymatora MNK może mieć znaczący wpływ na oczekiwany błąd prognozy. Ponadto powiększanie liczby użytych w modelu wskaźników indywidualnych może poprawiać dopasowanie modelu w próbie kosztem pogorszenia jakości prognoz poza próbą. W tym artykule sprawdzono, czy praktycznym rozwiązaniem problemu zbyt dużej liczby regresorów nie może być utworzenie na podstawie tych zmiennych jednej lub kilku zmiennych syntetycznych opisujących jak największą część zmienności szeregów w danej grupie. W badaniu empirycznym syntezy informacji dokonano za pomocą metod analizy czynnikowej. Jeżeli podejście to jest zasadne, to wartość predykcyjna syntetycznych wskaźników wyprzedzających powinna okazać się relatywnie większa niż wskaźników indywidualnych.

[6] W automatycznej procedurze TRAMO dla CPI ogółem znaleziono zaledwie dwie obserwacje odstające w okresie 2001–2012 (tj. sierpień 2000 roku i maj 2001 roku) wobec przynajmniej 9 tego typu wystąpień w takich komponentach CPI. Najwięcej w takich grupach dóbr jak: pieczywo i produkty zbożowe, nabiał, cukier, odzież, obuwie oraz w cenach usług edukacyjnych.

3. WSKAŹNIKI WYPRZEDZAJĄCE INFLACJI – PRZEGLĄD BADAŃ[7]

Metodę konstrukcji syntetycznych wskaźników wyprzedzających (ang. *leading indicators*) opracowano w amerykańskiej instytucji badawczej National Bureau of Economic Research (NBER) na potrzeby monitorowania i prognozowania stanu koniunktury ogólnogospodarczej (por. Burns i Mitchell, 1946). Wybierano w tym celu wskaźniki, które indywidualnie wykazywały największą zdolność do objaśnienia zmian przyszłej aktywności, a następnie nadawano im arbitralne wagi w mierniku syntetycznym. Usystematyzowaniem tej koncepcji zajęli się Stock i Watson (1989), którzy identyfikację wskaźników wyprzedzających oparli na zdolności prognozowania zmian szeregu referencyjnego (np. koniunktury czy inflacji). Przy czym przez szereg referencyjny rozumiano zmienną nieobserwowalną wyznaczaną na podstawie zmian wspólnych dla większości obserwowalnych aproksymant tego procesu[8]. Wspólne czynniki określane są „siłami sprawczymi" (ang. *driving forces*), a popularnym sposobem ich uzyskiwania są metody analizy czynnikowej wprowadzone do makroekonomii przez Geweke'a, Sargenta i Simsa w latach 70. XX wieku (za Stock i Watson, 2002a).

W latach 90. metody analizy czynnikowej przeprowadzanej na podstawie dużych zbiorów danych wykazały przydatność w prognozowaniu procesów makroekonomicznych. W 1998 roku Stock i Watson jako jedni z pierwszych[9] do wyznaczenia syntetycznego wskaźnika wyprzedzającego inflacji użyli metod analizy czynnikowej, a w 1999 r. zastosowali je do modelowania krzywej Philipsa. W następnych latach liczba badań wykorzystujących metody analizy czynnikowej w prognozowaniu inflacji szybko rosła, a wśród nich warto wymienić artykułu Angelini i in. (2001), Stocka i Watsona (2003), Banerjee i in. (2005). W ramach tych badań pojawiają się dyskusje na temat optymalnych rozmiarów zbiorów danych do analizy czynnikowej (Boivin, Ng, 2006), zgodnych metod estymacji modeli czynnikowych (por. Stock i Watson, 2002a, Forni i in., 2005, Doz i in., 2006), sposobów wykorzystania dynamicznego modelu czynnikowego w prognozowaniu (por. Boivin i Ng, 2005) oraz sposobu postępowania w sytuacji asynchronicznego napływu danych ekonomicznych (w polskiej literaturze por. Łupiński, 2012).

W dotychczasowych badaniach statystycznych nad wskaźnikami wyprzedzającymi inflacji analizowano najczęściej następujące grupy zmiennych: komponenty inflacji (Hendry i Hubrich, 2006; Kapetanios, 2004), oraz konstruowane na ich podstawie syntetyczne miary inflacji bazowej określane jako *core inflation* (Giannone i Matheson, 2007; Cristadoro i in., 2005), czy *pure inflation* (Reis i Watson, 2010, a dla Polski artykuł Brzozy-Brzeziny i Kotłowskiego, 2009), oczekiwania inflacyjne

---

[7] Ta część jest skróconą i uaktualnioną wersją przeglądu literatury zawartego w Szafrański (2011).

[8] Oprócz części wspólnej każdy szereg aproksymant szeregu referencyjnego zawiera zmiany swoiste zwane idiosynkratycznymi.

[9] Ostateczną wersję pracy opublikowano w 2002 roku (Stock i Watson, 2002b).

osób prywatnych i podmiotów gospodarczych (por. Ang i in., 2007, dla Polski Łyziak i Stanisławska, 2008), mierniki koniunktury stosowane w różnych wariantach krzywej Philipsa (por. ich przegląd w pracy Stocka i Watsona, 2008), miary podaży pieniądza (Hofmann, 2008), czy informacje z rynku finansowego o cenach i rentowności aktywów finansowych, a także czynnikach terminowej struktury stóp procentowych (Stock i Watson, 2003).

Badania dotyczące regularnego prognozowania inflacji w Polsce na podstawie metody wskaźników wyprzedzających nie są zbyt liczne. Na podstawie tradycyjnej metody NBER w ramach prac instytutu badawczego Bureau for Investment and Economic Cycles (BIEC) powstaje syntetyczny wskaźnik wyprzedzający dla inflacji w Polsce, tak zwany Wskaźnik Przyszłej Inflacji (WPI) – por. Białowolski i Żochowski (2006)[10]. Słabością tego podejścia jest opieranie się na ustalanych i zmienianych arbitralnie wagach. Z kolei podejścia, w których autorzy korzystają w sposób systematyczny z relatywnie dużych zbiorów danych można znaleźć w ramach badań porównawczych nad prognozowaniem zharmonizowanego indeksu inflacji (HICP) dla grupy 5 krajów (Polski, Czech, Węgier, Słowacji i Słowenii) wstępujących do Unii Europejskiej (UE) w 2004 roku – zob. Banerjee i in. (2004) i dla większej grupy nowych członków UE w artykule Arratibel i in. (2009). W tym pierwszym badaniu stosowano modele ze wspólnymi czynnikami uzyskanymi na podstawie kilkudziesięciu kwartalnych szeregów czasowych (56 dla Polski), a w tym drugim tradycyjne modele z jednym lub dwoma indywidualnymi wskaźnikami wyprzedzającymi wykazując ich wyższość nad prostymi metodami analizy szeregów czasowych.

W Polsce badania inflacji prowadzone na podstawie dynamicznych modeli czynnikowych i dużych zbiorów danych pojawiły się stosunkowo niedawno – por. artykuły Kotłowskiego (2008), Baranowskiego, i in. (2010). W pracach tych podobnie jak w tym opracowaniu do wyznaczenia wspólnych czynników wykorzystano metodę analizy głównych składowych (ang. *principal component analysis*, dalej PCA), której własności asymptotyczne (gdy liczba stacjonarnych szeregów czasowych oraz ich obserwacji po czasie rośnie do nieskończoności) opisano m.in. w Stock i Watson (2002a), Bai (2003). Warto wspomnieć, że w wyniku zastosowania metody PCA dochodzi do redukcji liczby zmiennych objaśniających w równaniu prognostycznym. Ze względu na potencjalną współliniowość niektórych regresorów redukcja ta powinna mieć wpływ na uzyskanie precyzyjniejszych ocen parametrów. Jednakże głównym celem zastosowania metody PCA w analizie czynnikowej dużych zbiorów danych jest wydobycie w sposób syntetyczny informacji o wspólnych czynnikach, która może być użyteczna w prognozowaniu inflacji.

W przeprowadzonym badaniu wstępnie ograniczono liczbę zmiennych przeznaczonych do analizy czynnikowej metodami statystycznymi. Redukcja ta nie jest jedy-

[10] Wskaźnik WPI zawiera jako komponenty następujące zmienne: kursy walutowe, indeksy cen importu i usług, zadłużenie sektora publicznego i osób prywatnych, jednostkowe koszty pracy, oraz dane ankietowe z testów koniunktury przemysłu przetwórczego IRG-SGH.

nie zabiegiem czysto praktycznym, ułatwiającym ekonomiczną interpretację wyników eksperymentu. Jest to również przyczynek do obecnej w literaturze dyskusji na temat statystycznych metod doboru zmiennych do analizy czynnikowej. Metody te, znane pod nazwą metod „najlepszych predyktorów" (tłum. własne ang. *targeted predictors*, termin użyty po raz pierwszy w artykule Bai i Ng, 2008), wskazują na korzyści z pominięcia w analizie czynnikowej szeregów o dużej zmienności idiosynkratycznej (swoistej, własnej). Ich obecność może bowiem znacząco obniżać dokładność, z jaką wyznacza się nieobserwowalne wspólne czynniki (por. Boivin i Ng, 2006).

4. EKSPERYMENT PROGNOSTYCZNY

4.1. DANE I ICH PRZYGOTOWANIE

Do przeprowadzenia badania wykorzystano bazę danych zawierającą te informacje makroekonomiczne i finansowe, które w okresie od lipca 2009 r. do grudnia 2012 r. były co miesiąc publicznie dostępne w dniu publikacji wskaźnika CPI przez GUS[11]. Daje to łącznie 42 generacje danych, w których najnowsze opublikowane dane często podlegały zmianom i rewizjom statystycznym, a niektóre z nich uzyskiwano na podstawie informacji za bieżący (niepełny) miesiąc[12]. W skład bazy wchodzą najbardziej aktualne obserwacje dotyczące realizacji 189 miesięcznych szeregów czasowych, z których najstarsze sięgają stycznia 2000 roku. Niektóre z szeregów charakteryzują się późniejszym terminem publikacji niż termin ogłaszania wskaźnika CPI, czyli tzw. opóźnieniem publikacyjnym, które sięga od 1 do kilku miesięcy. Inne, w tym dane z rynków finansowych, paliwowych i rolnych, dotyczą miesiąca następnego w stosunku do opublikowanego wskaźnika CPI. Przy przyjętej metodzie prognozowania (por. 4.2) jako zasadę przyjęto, że dane z okresu t, to dane dostępne w okresie t, a zatem nie zawsze opisują one sytuację gospodarczą w okresie t. W skład bazy wchodzą zmienne, które przyporządkowano do następujących, jednorodnych tematycznie grup danych:

- komponenty wskaźnika CPI (grupa CPI i ZYWU),
- czynniki dotyczące kreacji pieniądza (PIEN),
- czynniki determinujące koszty produkcji i dystrybucji dóbr (z podziałem na krajowe KRAJ i zagraniczne ZAGR),
- wskaźniki (tzw. „twarde dane" GUS) dotyczące sytuacji ogólnogospodarczej (popytowe, GUS),

[11] Inflacja za miesiąc poprzedni jest publikowana co do zasady około 15 dnia bieżącego miesiąca. Ze względu na układ dni wolnych w danym roku termin ten nie jest stały i podawany jest co roku przez GUS w kalendarium publikacji danych.

[12] Dotyczy oszacowań wskaźników średniomiesięcznych, które uzyskiwano na podstawie danych dostępnych w chwili prognozowania jedynie za pierwszą połowę miesiąca, a w szczególności danych dziennych z rynków finansowych oraz danych tygodniowych z rynku paliw i surowców rolnych.

– informacje dotyczące oczekiwań (w tym ankietowe wskaźniki koniunktury KON, oczekiwania inflacyjne osób prywatnych CiH, oraz wskaźniki dochodowości na rynku finansowym RF).

Zbiorcze informacje o zawartości bazy danych zawarto w Załączniku. Wszystkie potencjalne wskaźniki indywidualne, poddano rekurencyjnie procedurze odsezonowania (o ile metoda Tramo/Seats wskazała na obecność wahań sezonowych). Następnie wyrównane sezonowo szeregi poddano takim transformacjom, żeby w ich wyniku otrzymać szeregi stacjonarne. Szczegóły dotyczące źródeł, odsezonowania i transformacji danych przedstawiono w Załączniku.

4.2. PRZEDMIOT PROGNOZY I METODA PROGNOZOWANIA

Zmienną objaśnianą we wszystkich analizowanych modelach wskaźników wyprzedzających są wskaźniki inflacji CPI w ujęciu miesiąc do miesiąca (m/m). W badaniu wykorzystano metodę prognozowania bezpośredniego[13] (ang. *direct forecasting*), czyli dla prognozy wyznaczanej na h miesięcy naprzód (h = 1,...,12) konstruowany jest model prognostyczny wykorzystujący jedynie ostatnie dostępne obserwacje o zmiennych objaśniających (tj. sprzed h miesięcy w stosunku do ostatnio opublikowanej inflacji). Prognozy oparte na liniowych modelach wskaźników wyprzedzających (*LI*) przyjmują następującą postać ogólną:

$$\hat{\pi}_{t+h|t} = \sum_{i=1}^{I} \alpha_i^{(h)} LI_{it}^{(h)} + sez_t, \tag{1}$$

gdzie:

$\hat{\pi}_{t+h|t}$ – prognoza wskaźnika inflacji CPI m/m (okres poprzedni=1) na okres $t + h$ uzyskana na podstawie informacji dostępnej do okresu t,

$LI_{i,t}^{(h)}$ – „i”-ty wiodący (wyprzedzający) wskaźnik inflacji wyznaczany oddzielnie dla każdego z horyzontów prognoz (h = 1,...,12) wg metod opisanych w pkt 4.3 i 4.4,

$\alpha_i^{(h)}$ – ocena parametru uzyskana metodą najmniejszych kwadratów (MNK) dla wskaźnika indywidualnego lub wspólnego czynnika (i = 1,...,I),

sez_t – oznacza estymowane łącznie ze wskaźnikami wyprzedzającymi deterministyczne efekty sezonowe, co *de facto* sprowadza się do uzmiennienia wyrazu wolnego w każdym miesiącu.

[13] Z przyjęcia metody bezpośredniej w prognozowaniu wynika określenie osobnych wskaźników i osobnych modeli na każdy horyzont prognozy (zob. Stock i Watson, 2008).

4.3. INDYWIDUALNE WSKAŹNIKI WIODĄCE

W modelu wskaźników indywidualnych wszystkie odsezonowane zmienne z bazy danych potraktowano jako potencjalnych kandydatów na wskaźniki wiodące. Przyjęto oznaczenia $LI_{i,t}^{(h)} = \tilde{x}_{j(h),t}$, gdzie $j(h)$ oznacza kolejność uporządkowania indeksów zmiennych $j = 1,2,...,189$ dla horyzontu prognozy h ($h = 1,...,12$). Rangowania zmiennych w $j(h)$ dokonano „w próbie" na podstawie kryterium jakości dopasowania równania (1) z jedną zmienną objaśniającą ($I = 1$), oddzielnie dla każdego z horyzontów prognozy. Ze względu na potencjalną niestabilność wyboru najlepszych predyktorów dokonywanego na podstawie kolejnych generacji danych ranking utworzono na podstawie okresu testowego obejmującego 17 generacji danych, przy czym pierwsza generacja zawiera dane dostępne w lipcu 2009 r., a ostatnia w listopadzie 2010 r. W ten sposób zmiennym przypisano średnie pozycje w rankingu, które następnie uporządkowano do postaci rankingu wskaźników $j(h)$ wyznaczonego oddzielnie na każdy horyzont prognozy. W tabeli 2 widoczne są trzy pierwsze pozycje każdego z tych rankingów, a syntetyczne informacje o pełnym rankingu 20 najlepszych wskaźników zaprezentowano w podziale na grupy zmiennych w tabeli 3. W celu sprawdzenia zdolności prognostycznej wybranych modeli „poza próbą" do modelu prognostycznego (1) włączane są kolejne zmienne od pozycji 1 do pozycji 30 określonej na podstawie tego rankingu.

Tabela 2.

Rankingi 3 „najlepszych w próbie" indywidualnych wskaźników wyprzedzających na horyzonty od 1 do 12 miesięcy (na podstawie generacji danych: lipiec 2009 – grudzień 2010 r.)

Rankingi	Pozycja 1		Pozycja 2		Pozycja 3	
Horyzonty prognoz	Nazwa grupy / symbol zmiennej	średnia pozycja	Nazwa grupy / symbol zmiennej	średnia pozycja	Nazwa grupy / symbol zmiennej	średnia pozycja
h=1	CPI bazowa 15%	1,0	CPI Wyposaż. mieszkań	2,4	CPI ogółem (SA)	3,8
h=2	CPI bazowa 15%	1,4	CiH IPSOS3	2,2	KRAJ Mleko MRol	5,1
h=3	CiH IPSOS4	2	CiH IPSOS1	4,0	CiH IPSOS3	4,1
h=4	CiH IPSOS1	1,6	CiH IPSOS4	2,0	CiH IPSOS3	4,7
h=5	CiH IPSOS3	1,4	CiH IPSOS1	3,3	CPI Opał	5,6
h=6	CiH IPSOS3	1,8	CiH IPSOS4	4,8	CPI Usługi medyczne	6,9

h=7	ZYWU Nabiał	1,5	RF Stopa 12M	2,3	RF Stopa 3M	4,3
h=8	ZYWU Nabiał	1,0	RF Stopa 12M	2,0	RF Stopa 2Y	3,1
h=9	RF Stopa 12M	1,1	RF Stopa 3M	2,5	RF Stopa 2Y	3,2
h=10	RF Stopa 10Y	1,1	ZYWU Napoje bezalk.	2,5	RF Stopa 2Y	2,6
h=11	ZYWU Napoje alk.	1,1	ZAGR MFW oliwa	2,8	RF WIG Telekom	2,9
h=12	WWG Bezrob. rejestr.	1,3	RF WIG Telekom	4,2	RF WIG 20	4,9

Średni ranking ustalono na podstawie rankingów dla 18 generacji danych w okresie testowym utworzonych w oparciu o jakość dopasowania modelu (1) dla wskaźnika indywidualnego (). Oznaczenia niektórych zmiennych; w grupie wskaźników CPI i ZYWU (zawierającej wszystkie odsezonowane i w ujęciu m/m); „bazowa 15%" – inflacja bazowa „15% średnia obcięta", ogółem (SA) – wyrównany sezonowo wskaźnik CPI ogółem, pozostałe grupy CPI zgodne z opisem tworzone na podstawie jednej lub kilku grup COICOP, w grupie CiH; IPSOSx – udziały odpowiedzi na pytanie „x" w ankiecie dot. oczekiwań inflacyjnych osób prywatnych IPSOS, w grupie KRAJ; Mleko MRol– mleko spożywcze pasteryzowane (źródło: tygodniowy monitoring rynków rolnych Ministerstwa Rolnictwa i Rozwoju Wsi, dalej MRol), ZAGR; MFW – ceny surowców na rynkach światowych, RF; stopy dochodowości na rynku finansowym (na wybrane terminy), oraz dochodowości indeksów giełdowych (WIG). Oznaczenia nazw grup zmiennych i źródło danych zob. Załącznik. Źródło: obliczenia własne.

Tabela 3.

Udział procentowy zmiennych pochodzących z grup prognostycznych w rankingu 20 wskaźników indywidualnych najsilniej skorelowanych z przyszłą inflacją w horyzoncie od 1 do 12 miesięcy

Grupy	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9	h=10	h=11	h=12
CiH	15	15	15	20	20	15	15	0	5	5	5	0
PiB	0	15	5	0	0	0	0	5	0	5	5	10
RF	0	5	0	0	10	10	40	35	35	25	35	30
KRAJ	30	20	30	20	5	5	0	15	15	0	10	10
ZAGR	0	0	0	5	5	25	0	0	0	20	30	15
WWG	0	0	0	0	0	0	10	5	5	10	5	10
ZWG	0	0	0	0	0	5	0	0	0	0	0	0
PIEN	0	5	20	20	30	20	15	0	5	10	5	0
CPI	40	15	10	15	20	15	10	15	15	15	0	20
ZYWN	15	25	20	20	10	5	10	25	20	10	5	5

Źródło: obliczenia własne.

4.4. SYNTETYCZNE WSKAŹNIKI WIODĄCE

Wskaźniki syntetyczne gromadzą informację zawartą w zmiennych występujących w rankingu na pozycji 1 do R (R = 5,10,15,20,25,30) skondensowano do jednego lub kilku wspólnych czynników za pomocą metody PCA:

$$LI_{i,t}^{(h)} = \sum_{j=1}^{R} \hat{\lambda}_j^{(h)} \tilde{x}_{j(h),t}, \tag{2}$$

gdzie $\hat{\lambda}_j$ jest współczynnikiem stojącym przy zmiennej o numerze i w wektorze własnym odpowiadającym j-tej największej wartości własnej macierzy kowariancji wystandaryzowanych zmiennych $\tilde{x}_{j(h),t}$. Wspólne czynniki wyznaczone na podstawie formuły (2), oddzielnie na każdy horyzont h = 1,2,...,12,, wykorzystywane są następnie jako wskaźniki wiodące w równaniu (1).

Uwaga: w modelu indywidualnych wskaźników wyprzedzających uwzględniono następujące warianty dotyczące liczby wskaźników wyprzedzających I = 1,2,3,5,10,15,20,25,30. Stąd dla tych modeli na rysunku 2 i rysunku 3 oraz w tabeli 4 przyjęto oznaczenia: 1 zm, 2 zm, ..., 30 zm. Natomiast w modelu syntetycznych wskaźników wyprzedzających parametr I = 1,2,3 oznacza liczbę uwzględnionych w modelu wspólnych czynników uzyskanych metodą analizy głównych składowych spośród 5, 10, 15 lub 20 zm (np. 2 PCA-5 oznacza model z dwoma wspólnymi czynnikami wyznaczony na podstawie 5 najlepszych predyktorów).

4.5. ZAŁOŻENIA EKSPERYMENTU PROGNOSTYCZNEGO

Eksperyment polegał na wykonaniu prognoz inflacji na h miesięcy naprzód[14] (h = 1,2,...,12) każdorazowo na podstawie oddzielnej generacji danych dostępnych co miesiąc w okresie od grudnia 2010 – do listopada 2012 r. Ze względu na to, że przedmiotem zainteresowania podmiotów gospodarczych jest tempo zmian cen w ujęciu rok do roku (r/r) wyniki prognoz punktowych zaprezentowano po ich przeliczeniu na wskaźniki r/r. Prognozy inflacji r/r są następnie porównane z realizacjami za pomocą syntetycznych miar błędów prognoz RMSE, ME i MAE, tj. odpowiednio pierwiastka błędu średniokwadratowego, błędu średniego i błędu absolutnego. Parametry modeli prognostycznych (1) reestymowano metodą najmniejszych kwadratów z wykorzystaniem jedynie danych dostępnych do okresu t oraz stałego okna estymacji (ang. *rolling*

[14] Uwaga: prognozowanie na jeden okres naprzód (h = 1), zwane *nowcastingiem*, dotyczy prognozy inflacji na bieżący miesiąc. Określenia okresów naprzód użyto jedynie w kontekście dostępności danych, a nie w odniesieniu do bieżącego okresu, w którym te dane są publikowane.

window)[15]. Przyjęty w eksperymencie dla wszystkich metod układ badania jakości prognoz i dostępna baza danych pozwala na weryfikację prognoz *ex post* inflacji rok do roku kolejno na podstawie: 24 obserwacji dla prognoz na $h = 1$, dwudziestu dwóch – na $h = 2$, …, i dwunastu – na $h = 12$.

5. WYNIKI EMPIRYCZNE

Jako podstawowe kryterium dokładności prognozy punktowej, w opisanym w części 4.5 okresie weryfikacji obejmującym lata 2011–2012, przyjęto pierwiastek błędu średniokwadratowego (RMSE) liczony oddzielnie dla każdego horyzontu prognoz. W celu porównania prognoz z różnych modeli liczony jest również relatywny spadek błędu RMSE (oznaczający poprawę jakości prognoz) w stosunku do prognozy bazowej wyrażony w %. Pomocniczo analizowano również obciążenie prognoz za pomocą miary ME i relacji ME/MAE (dalej określanej jako stopień obciążenia prognoz[16]). Jakość prognoz sprawdzono pod kątem wpływu na nią liczby wiodących wskaźników indywidualnych i liczby wspólnych czynników użytych w równaniu (1), oraz obydwu tych elementów jednocześnie.

5.1. MODELE WSKAŹNIKÓW INDYWIDUALNYCH W PRÓBIE

Ranking zmiennych najsilniej skorelowanych z przyszłą inflacją w horyzoncie od 1 do 12 miesięcy naprzód okazał się być relatywnie stabilny w kolejnych 17 generacjach danych poprzedzających okres weryfikacji prognoz. Zmienne najsilniej skorelowane pozostawały na najwyższych pozycjach rankingu przez kilkanaście miesięcy[17], na co wskazują średnie pozycje z kolejnych miesięcy dla 3 pierwszych zmiennych – por. tabela 2. Analizując zawartość ekonomiczną zmiennych dla poszczególnych horyzontów prognoz na pozycjach 1–3 w rankingu (zob. tabela 2) i grup zmiennych w rankingu 20 zmiennych najsilniej skorelowanych z przyszłą inflacją (zob. tabela 3) można wyciągnąć następujące wnioski:

1. W prognozie wskaźnika inflacji za bieżący miesiąc tj. dla h=1 w modelu (1) najbardziej przydatne są ostatnie informacje pochodzące ze wskaźników inflacji bazowej, inflacji CPI i jej nieżywnościowych komponentów, co może być oznaką

[15] Stałe okno estymacji zastosowano między innymi w celu złagodzenia konsekwencji potencjalnie zmiennego wzorca sezonowości i wpływu na wyniki estymacji obserwacji pochodzących z początku próby.

[16] Relacja tych miar błędów jest zawsze z przedziału (-1,1). Jej graniczne wartości świadczą o systematycznym przeszacowaniu (dolna granica przedziału) lub systematycznym niedoszacowaniu (górna granica) zmiennej prognozowanej.

[17] Również na kolejnych pozycjach wybór zmiennych wykazywał się dużą stabilnością. Przykładowo wszystkie wskaźniki indywidualne do 7 pozycji w rankingu uzyskiwały w kolejnych miesiącach pozycję nie gorszą niż 10, a do 10 pozycji – nie gorszą niż 15.

uporczywości (ang. *persistence*) inflacji rozumianej jako tendencja do utrzymywania się efektów impulsowego szoku dotyczącego cen w kolejnym miesiącu (por. Hertel, Leszczyńska 2013). W rankingu 20 najsilniej skorelowanych wskaźników indywidualnych dla *nowcastingu* indeks CPI i jego komponenty stanowią ponad 55% wybranych zmiennych, a 40% po wyłączeniu żywnościowych komponentów CPI – grupa ZYWU. Ważną grupą prognostyczną (30% udział) są również wskaźniki związane ze zmianą kosztów dóbr konsumpcyjnych na rynku krajowym (KRAJ).

2. Dla prognoz od 2 do 6 miesięcy naprzód na najwyższych pozycjach w rankingu dominują informacje pochodzące z badań ankietowych konsumentów dotyczące ich oczekiwań inflacyjnych (IPSOS). Wśród wskaźników na 20 najwyższych pozycjach stanowią one (wraz z badaniami ankietowymi cen i handlu – grupa CiH) od 15% do 20%. W rankingu 20 wskaźników występują ponadto komponenty CPI, choć ich udział maleje wraz z horyzontem prognoz (od 40% dla h=2 do 20% dla h=5 i h=6). Podobnie wraz z wydłużeniem horyzontu prognozy tracą na znaczeniu informacje o kosztach krajowych (KRAJ np. informacje o cenach skupu produktów rolnych), a wzrasta znaczenie danych dotyczących podaży pieniądza (PIEN).
3. Dla prognoz od 7 do 12 miesięcy naprzód najliczniej występującą grupą wskaźników indywidualnych (od 25% do 40% wskaźników) są dane pochodzące z rynku finansowego (RF, w tym przede wszystkim stopy procentowe z rynku finansowego, a na 11 i 12 miesięcy również stopy dochodowości rynku akcji). Kolejne grupy zmiennych w rankingu dotyczą kosztów dóbr konsumpcyjnych wytwarzanych w kraju (KRAJ) i zagranicą (ZAGR). W tej grupie prognoz koszty dóbr importowanych mają znaczenie tylko dla prognoz na 10–12 miesięcy naprzód. Wśród pozostałych wskaźników wyprzedzających w rankingu występują też komponenty CPI nieżywnościowe (CPI) i te związane z żywnością i napojami (ZYWU). Udział tej drugiej grupy wyraźnie maleje wraz z wydłużeniem horyzontu prognozy.
4. O wiele mniejsze znaczenie prognostyczne w prognozach na okresy od 7 do 12 miesięcy wykazują „twarde" dane GUS na temat sytuacji gospodarczej branż i rynku pracy (WWG) i prognozowanej koniunktury gospodarczej (PiB). Wartość predykcyjna informacji o kondycji handlu zagranicznego (ZWG) w konstrukcji wskaźników wyprzedzających jest najmniejsza spośród wszystkich analizowanych grup zmiennych.

5.2. MODELE WSKAŹNIKÓW INDYWIDUALNYCH A MODELE WSKAŹNIKÓW SYNTETYCZNYCH – WŁASNOŚCI PROGNOSTYCZNE

Modelem bazowym do porównań modeli wskaźników indywidualnych i syntetycznych uczyniono model (1) z jednym wskaźnikiem wiodącym (oddzielnie na każdy horyzont). Na rysunku 2 przedstawiono poprawę jakości prognoz (spadek RMSE) dla modeli wskaźników indywidualnych i syntetycznych w wybranych (reprezentatywnych) horyzontach prognoz (h=1,3,6,9,12) oraz średnią poprawę jakości prognoz dla wszystkich 12 horyzontów w stosunku do modelu z jedną zmienną. W okresie weryfikacji skonstruowanym na podstawie generacji danych od grudnia 2010 r. do listopada 2011 r. można sformułować następujące wnioski:

1. Jakość prognoz dla wskaźników indywidualnych początkowo rośnie wraz ze wzrostem liczby wskaźników wiodących (chociaż nie monotonicznie) osiągając dla 5 najlepszych w próbie wskaźników wiodących (5 zm) RMSE niższe w stosunku do modelu bazowego od 2% dla h = 10 do 17% dla h = 1 (tj. średnio o mniej niż 10%). Gdy liczba zmiennych w modelu wskaźników indywidualnych przekracza 20, to jakość prognoz stopniowo się pogarsza w stosunku do prognoz z jednym wskaźnikiem, szczególnie dla horyzontów prognoz pow. 5 miesięcy.

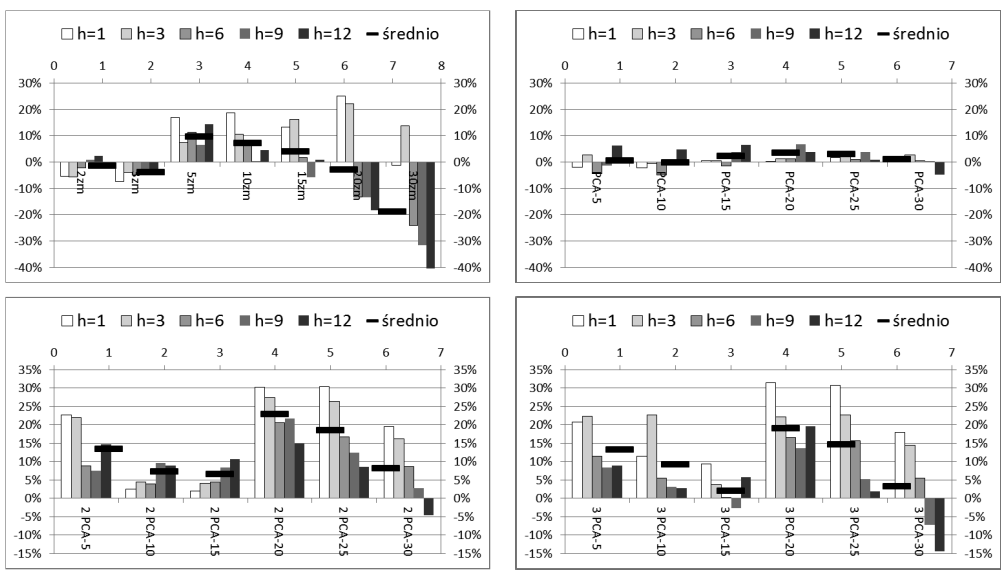
2. Kondensowanie informacji ze wskaźników indywidualnych do jednego wspólnego czynnika przynosi skromną poprawę jakości prognoz średnio poniżej 5% w analizie czynnikowej dla 20 zmiennych (1-PCA-20), chociaż wraz ze wzrostem liczby zmiennych relatywne pogorszenie jakości prognoz nie jest tak znaczne jak dla wskaźników indywidualnych.

3. Włączanie do modelu wskaźników wyprzedzających drugiego (2-PCA) i trzeciego (3-PCA) wspólnego czynnika (tj. kolejnych głównych składowych macierzy korelacji uzyskanych metodą PCA) przynosi znaczną poprawę jakości prognoz, największą dla około 20 wskaźników wyprzedzających użytych w analizie czynnikowej – średnio o 23% dla 2 czynników (od 11% na h = 11 do 37% na h = 2), oraz o 19% dla 3 wspólnych czynników. Dalsze zwiększanie liczby zmiennych do analizy czynnikowej skutkuje spadkiem korzyści ze stosowania analizy czynnikowej, szczególnie dla horyzontów prognoz powyżej 6 miesięcy.

4. Porównanie jakości prognoz wybranych modeli wskaźników indywidualnych i wybranego modelu wskaźników syntetycznych dla 20 zmiennych i 2 wspólnych czynników wskazuje na korzyści ze stosowania analizy czynnikowej dla prognoz od 4 do 11 miesięcy wprzód (por. tabela 4 i rysunek 3). Na podstawie prezentowanej metody wskaźników syntetycznych uzyskano w dwuletnim okresie weryfikacji prognozy przynajmniej o porównywalnej precyzji i obciążeniu w stosunku do pro-

gnoz korzystających z dowolnego zestawu optymalnie dobieranych wskaźników indywidualnych.

5. Kondensowanie informacji w postaci 2–3 wspólnych czynników zwiększa odporność prognoz opartych na wskaźnikach wyprzedzających na zmianę liczby zmiennych w modelu w porównaniu z metodami opartymi na wskaźnikach indywidualnych. Ma to praktyczne znaczenie, gdyż na ogół nie znamy optymalnej liczby zmiennych, których należy użyć do konstrukcji wskaźników wyprzedzających. Wówczas synteza informacji w postaci metod analizy czynnikowej jest bezpiecznym, choć nie zawsze najbardziej efektywnym sposobem uzyskiwania prognoz na podstawie dużych zbiorów danych.



Rysunek 2. Relatywna jakość prognoz w okresie weryfikacji od grudnia 2010 r. do listopada 2012 r. Jakość prognoz policzono dla modeli indywidualnych i syntetycznych wskaźników wyprzedzających w porównaniu z wariantem bazowym (tj. modelem z jednym wskaźnikiem indywidualnym na każdy horyzont) i wyrażono jako procentowy spadek RMSE. Określenie „średnia" dotyczy średniej jakości prognoz dla wszystkich horyzontów.

Źródło: obliczenia własne.

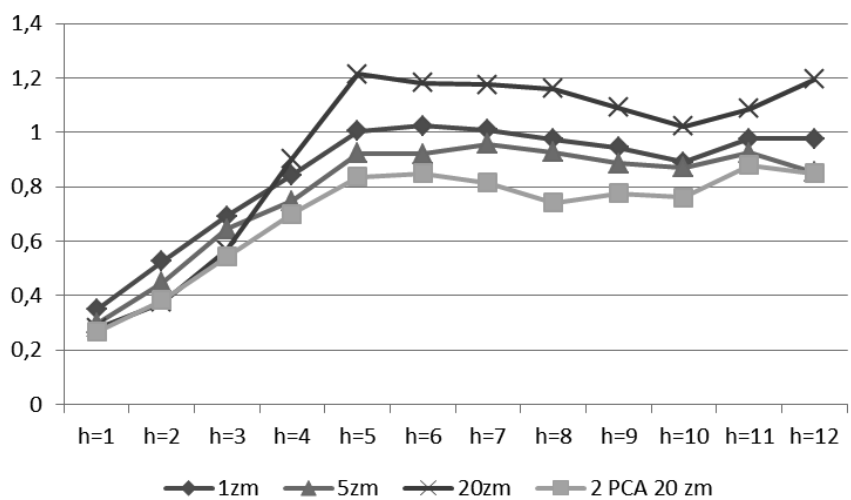
Tabela 4.

Błędy prognoz poza próbę dla wybranych modeli wskaźników wyprzedzających i horyzontów w okresie weryfikacji od stycznia 2011 r. do grudnia 2012 r.

Model	Miara błędów					
1 zm	RMSE	0,35	0,69	1,02	0,94	0,98
	MAE	0,28	0,61	0,85	0,77	0,85
	ME	-0,07	-0,08	-0,19	-0,36	-0,68
5 zm	RMSE	0,30	0,64	0,92	0,89	0,85
	MAE	0,24	0,50	0,76	0,74	0,75
	ME	-0,04	-0,13	-0,13	-0,06	-0,37
20 zm	RMSE	0,28	0,57	1,18	1,09	1,19
	MAE	0,22	0,44	0,97	0,89	1,09
	ME	0,03	-0,04	-0,03	0,07	-0,23
2 PCA 20 zm	RMSE	0,27	0,54	0,85	0,78	0,85
	MAE	0,22	0,46	0,68	0,62	0,72
	ME	0,01	-0,01	-0,06	-0,10	-0,33

Oznaczenia modeli dotyczą modeli z jedną (1 zm), pięcioma (5 zm) i dwudziestoma (20 zm) wskaźnikami indywidualnymi, oraz modelu wskaźników syntetycznych na podstawie 20 zmiennych i dwóch wspólnych czynników (2 PCA-20 zm).

Źródło: obliczenia własne.



Rysunek 3. Błędy RMSE (w pp.) dla wybranych modeli wskaźników wyprzedzających w okresie od stycznia 2011 do grudnia 2012 r. Oznaczenia modeli jak w Tabeli 4.
Źródło: obliczenia własne.

6. PODSUMOWANIE

Wnioski, jakie płyną z analizy modeli wskaźników wiodących dla inflacji w Polsce można wstępnie podsumować w następujący sposób. Po pierwsze, metoda wyboru zmiennych na podstawie analizy siły ich korelacji z przyszłą inflacją w modelu wskaźników wyprzedzających opartym o zasadę prognozowania bezpośredniego i uwzględniającym sezonowość daje (w okresie testowym) stosunkowo stabilne rezultaty z punktu widzenia wyboru predyktorów. Spośród zmiennych najsilniej skorelowanych z przyszłą inflacją dla horyzontów prognoz do 6 miesięcy występują zmienne niosące informację o krótkookresowej uporczywości inflacji (tj. miary agregatów CPI i jego komponentów), oraz oczekiwania inflacyjne gospodarstw domowych. Dla dłuższych horyzontów prognoz zyskują na znaczeniu informacje pochodzące z rynku finansowego, a także komponenty CPI dotyczące żywności. Po drugie, początkowo wraz ze wzrostem liczby indywidualnych wskaźników wyprzedzających w modelu prognostycznym rośnie jakość prognoz modelowych, co może świadczyć o tym, że wiele czynników kształtuje przyszłą inflację. Po trzecie, potwierdzony został wniosek z artykułu Stocka i Watsona (2006) o pogarszaniu się własności prognostycznych modelu wskaźników indywidualnych, gdy liczba zmiennych objaśniających w modelu prognostycznym przekroczy pewien pułap. Wówczas, jak pokazują wyniki wielu analiz, remedium na utratę dokładności estymatora może być syntetyzowanie informacji pochodzących z wielu źródeł za pomocą metod analizy czynnikowej i statystycznego doboru zmiennych do tej analizy. Dla prognozowania inflacji w Polsce w okresie weryfikacji obejmującym lata 2011–2012 daje ona szczególnie dobre rezultaty w postaci redukcji błędów *ex post* dla liczby wspólnych czynników większej niż jeden. Podobnie jak w modelu dla wskaźników indywidualnych również w modelu czynnikowym dołączanie do modelu kolejnych zmiennych (coraz słabiej skorelowanych w próbie) ma wpływ na pogorszenie jakości prognoz, ale skutki nie są tak znaczące jak w modelach wskaźników indywidualnych.

Uniwersytet Łódzki oraz Instytut Ekonomiczny NBP

LITERATURA

[1] Ang, A., Bekaert, G., Wei, M., (2007), Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better? *Journal of Monetary Economics*, 54 (4), 1163–1212.

[2] Angelini, E., Henry, J., Mestre, R., (2001), Diffusion Index-based Inflation Forecasts for the Euro Area. Working Paper Series 061, European Central Bank.

[3] Arratibel, O., Kamps, C., Leiner-Killinger, N., (2009), Inflation Forecasting in the New EU Member States. Working Paper Series 1015, European Central Bank.

[4] Bai, J., (2003), Inferential Theory for Factor Models of Large Dimensions, *Econometrica*, Elsevier, 71 (1), 135–172.

[5] Bai, J., Ng, S., (2008), Forecasting Economic Time Series Using Targeted Predictors, *Journal of Econometrics*, Elsevier, 146 (2), 304–317.
[6] Banerjee, A., Marcellino, M., Masten, I., (2004), Forecasting Macroeconomic Variables for the Acceding Countries. Working Papers 260, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.
[7] Banerjee, A., Marcellino, M., Masten, I., (2005), Leading Indicators for Euro-Area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics*, 67 (s1), 785–813.
[8] Baranowski, P., Leszczyńska, A., Szafrański, G., (2010), Krótkookresowe prognozowanie inflacji z użyciem modeli czynnikowych, *Bank i Kredyt*, 41 (4), 23–44.
[9] Białowolski, P., Żochowski, D., (2006), Analiza własności prognostycznych komponentów WPI, w: Drozdowicz-Bieć M., (red.), *Wskaźniki wyprzedzające.* Prace i Materiały Instytutu Rozwoju Gospodarczego SGH. Warszawa.
[10] Białowolski, P., Żochowski, D., (2006), *Wskaźniki wyprzedzające*. Prace i Materiały Instytutu Rozwoju Gospodarczego SGH, Rozdz. w: *Analiza własności prognostycznych komponentów WPI*, 167–197.
[11] Boivin, J., Ng, S., (2005), Understanding and Comparing Factor-Based Forecasts, *International Journal of Central Banking*, 1 (3), 117–151.
[12] Boivin, J., Ng, S., (2006), Are More Data Always Better for Factor Analysis?, *Journal of Econometrics*, Elsevier, 132(1), 169–194.
[13] Burns, A. F., Mitchell, W. C., (1946), *Measuring Business Cycles*, Nowy Jork, NBER.
[14] Brzoza-Brzezina, M., Kotłowski, J., (2009), Bezwzględna stopa inflacji w gospodarce polskiej, *Gospodarka Narodowa*, 9 (37), 1–21.
[15] Cristadoro, R., Forni, M., Reichlin, L., Veronese, G., (2005), A Core Inflation Indicator for the Euro Area, *Journal of Money, Credit and Banking*, 37 (3), 539–560.
[16] Doz, C., Giannone, D., Reichlin, L., (2006), A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models, Working Paper Series 674, European Central Bank.
[17] Forni, M., Hallin, M., Lippi, M., Reichlin, L., (2005), The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting, *Journal of the American Statistical Association*, 100, 830–840.
[18] Giannone, D., Matheson, T., (2007), A New Core Inflation Indicator for New Zealand, CEPR Discussion Papers 6469.
[19] Hendry, D. F., Hubrich, K., (2006), Forecasting Economic Aggregates by Disaggregates, Working Paper Series 589, European Central Bank.
[20] Hertel, K., Leszczyńska, A., (2013), Uporczywość inflacji i jej komponentów – badanie empiryczne dla Polski, *Przegląd Statystyczny*, 60 (2), 187–210.
[21] Hofmann, B., (2008), Do Monetary Indicators Lead Euro Area inflation?, Working Paper Series 867, European Central Bank.
[22] Kapetanios, G., (2004), A Note on Modelling Core Inflation for the UK Using a New Dynamic Factor Estimation Method and a Large Disaggregated Price Index Dataset, *Economics Letters*, 85 (1), 63–69.
[23] Kotłowski, J., (2008), Forecasting Inflation with Dynamic Factor Model – the Case of Poland, Working Papers 24, Department of Applied Econometrics, Warsaw School of Economics.
[24] Łupiński M., (2012), Short-term Forecasting and Composite Indicators Construction with Help of Dynamic Factor Models Handling Mixed Frequencies Data with Ragged Edges, *Przegląd Statystyczny*, 59 (1), 48–73.
[25] Łyziak, T., Stanisławska, E., (2008), Consumer Inflation Expectations in Europe: Some Cross-Country Comparisons, *Bank i Kredyt*, 39 (9), 14–28.
[26] Reis, R., Watson, M. W., (2010), Relative Goods' Prices, Pure Inflation, and the Phillips Correlation, *American Economic Journal: Macroeconomics*, 2 (3), 128–157.

[27] Stock, J. H., Watson, M. W., (1989), New Indexes of Coincident and Leading Economic Indicators, w: Blanchard, O. J., Fischer, S., (red.), NBER Macroeconomics Annual 1989, 4, NBER Chapters, National Bureau of Economic Research, Inc, 351–409.
[28] Stock, J. H., Watson, M. W., (2002a), Forecasting Using Principal Components From a Large Number of Predictors, *Journal of the American Statistical Association,* 97 (460), 1167–1179.
[29] Stock, J. H., Watson, M. W., (2002b), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics*, 20 (2), 147–162.
[30] Stock, J. H., Watson, M. W., (2003), Forecasting Output and Inflation: The Role of Asset Prices, *Journal of Economic Literature*, 41 (3), 788–829.
[31] Stock, J. H., Watson, M. W., (2006), Forecasting with Many Predictors, w: Elliot G., Granger C.W.J., Timmermann A., (red.), *Handbook of Economic Forecasting*, Elsevier, Roz. 10 (1), 515–554.
[32] Stock, J. H., Watson, M. W., (2008), Phillips Curve Inflation Forecasts, NBER Working Papers 14322, National Bureau of Economic Research, Inc.
[33] Szafrański G., (2011), Krótkoterminowe prognozy polskiej inflacji w oparciu o wskaźniki wyprzedzające, Materiały i Studia, Zeszyt nr 263, NBP, Warszawa.

ZAŁĄCZNIK

ZAWARTOŚĆ MERYTORYCZNA I STATYSTYCZNA BAZY DANYCH

Tabela 5.

Informacje na temat zawartości zbioru danych

Symbol grupy	Zawartość grupy	Źródło danych	Liczba zmiennych	Opóźnienie publikacyjne*	Transformacje/ sezonowość (SA)
Komponenty wskaźnika CPI					
CPI	Wskaźnik CPI ogółem, inflacja bazowa i komponenty CPI bez grupy ZYWU	GUS	24	1 mies.	Indeksy m/m (SA)
ZYWU	Komponenty CPI grupy: żywność, napoje, używki	GUS	13	1 mies.	Indeksy m/m (SA)
Podaż pieniądza i czynniki jego kreacji					
PIEN	Podaż pieniądza, rezerwa obowiązkowa i składniki bilansu MIF	NBP	20	2–3 mies.	Tempa zmian m/m (SA)
Koszty produkcji i surowców w podziale na źródła krajowe i zagraniczne					
KRAJ	Koszty produkcji (PPI)	GUS	9	2 mies.	Indeksy m/m (SA)
	Koszty surowców rolnych	MRol	6	0,5 mies.	Tempa zmian m/m (SA)
	Ceny paliw i ich koszty (w tym akcyza)	e-petrol.pl	6	0,5 mies.	
ZAGR	Kursy walutowe PLN	NBP	4	0,5 mies.	Zmiany m/m
	Ceny dóbr i surowców (w PLN)	MFW	15	1 mies.	Tempa zmian m/m (SA)
Wskaźniki makroekonomiczne dot. bieżącej sytuacji ogólnogospodarczej					
MAKR	Wskaźniki aktywności w branżach	GUS	6	2 mies.	Indeksy m/m (SA)
	Budżet państwa	GUS	4	2 mies.	Tempa zmian m/m (SA)
	Dane z rynku pracy	GUS	10	2 mies.	Zmiany m/m (SA)
	Handel zagraniczny	GUS	10	3 mies.	Tempa zmian m/m (SA)
Dane dotyczące oczekiwań (w tym dane ankietowe i dane z rynku finansowego)					

CiH	Wskaźniki koniunktury (ceny i handel) Oczekiwania inflacyjne	GUS IPSOS	17 5	1 mies. 1 mies.	Zmiany m/m (SA)
KON	Wskaźniki koniunktury (ogółem, przemysł i budownictwo	GUS	25	1 mies.	Zmiany m/m (SA)
RF	Stopy procentowe instrumentów dłużnych (lokaty mbank., obligacje SP) Stopy zwrotu (indeksy giełdowe)	Reuters Reuters	7 8	0,5 mies. 0,5 mies.	Tempa zmian m/m

* Opóźnienie publikacyjne liczone jest oddzielnie dla danych pochodzących z generacji danych. Jest to różnica między końcem miesiąca sporządzenia prognozy a okresem, którego dotyczą najnowsze dane dostępne w danej generacji danych, np. opóźnienie publikacyjne 1 miesiąc oznacza, że w miesiącu sporządzenia prognozy dostępne są dane za poprzedni miesiąc, „0,5" miesiąca oznacza, że do obliczenia średniej za bieżący miesiąc wykorzystywane są dane dzienne lub tygodniowe dot. pierwszej połowy bieżącego miesiąca.

Źródło: opracowanie własne.

SYNTETYCZNE WSKAŹNIKI WYPRZEDZAJĄCE W PROGNOZOWANIU INFLACJI W POLSCE

Streszczenie

W artykule przedstawiono wyniki procedury doboru zmiennych pochodzących z dużego zbioru danych ekonomicznych i finansowych do systematycznego, krótkookresowego prognozowania wskaźnika inflacji (CPI) w Polsce. Doboru zmiennych najsilniej skorelowanych z przyszłą inflacją w horyzoncie od 1 do 12 miesięcy naprzód, zwanych wskaźnikami wyprzedzającymi inflacji, dokonano na podstawie generacji danych rzeczywiście dostępnych co miesiąc od lipca 2009 do listopada 2010 roku (okres testowy). Następnie jakość prognoz w oparciu o modele wskaźników indywidualnych porównano ze wskaźnikami syntetycznymi w okresie od grudnia 2010 do listopada 2012 roku (okres weryfikacji) stosując bezpośrednią metodę prognozowania. W porównaniu do modeli wskaźników indywidualnych ustalonych na podstawie rankingu w okresie testowym wskaźniki syntetyczne uzyskane za pomocą metod analizy czynnikowej okazują się być dobrymi predyktorami przyszłej inflacji pod względem precyzji, nieobciążoności i odporności prognoz na zmiany liczby wskaźników i wspólnych czynników.

Słowa kluczowe: wskaźniki wyprzedzające, analiza czynnikowa, inflacja w Polsce

COMPOSITE LEADING INDICATORS IN FORECASTING INFLATION IN POLAND

Abstract

We present the procedure for a systematic selection of short-term leading indicators for Polish consumer inflation (CPI) from large number of predictors. Using data set of 189 economic and financial time series (available from July 2009 to November 2012) we select the indicators highly (in-sample) correlated with future inflation from 1 to 12 months ahead. Then we perform a direct forecasting exercise on a real-time data base from December 2010 to November 2012. Compared to individual leading indicators we find that few common factors extracted from a principal component analysis (named composite leading indicators) are successful tools for predicting future inflation being precise, unbiased, and more robust to changes in a number of predictors and common factors.

Keywords: leading indicators, factor analysis, inflation in Poland

PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

POLSCY STATYSTYCY I MATEMATYCY

JUBILEUSZ 70-LECIA URODZIN PROFESORA WŁADYSŁAWA MILO SYLWETKA I DOROBEK NAUKOWY

Władysław Milo urodził się 9 czerwca 1943 r. w Majdanie Lipowieckim (pow. Lubaczów). Liceum Ogólnokształcące ukończył w roku 1960 w Lubaczowie. W latach 1961-1966 studiował na Wydziale Ekonomii Politycznej Uniwersytetu Warszawskiego. Tytuł magistra ekonomii uzyskał za pracę pt. „Zastosowania programowania liniowego w optymalizacji budownictwa" w czerwcu 1966 r.

W latach 1965-1966 pracował w Instytucie Organizacji i Mechanizacji Budownictwa w Warszawie, a następnie w październiku 1966 r. został zatrudniony w Uniwersytecie Łódzkim, w którym pracuje do chwili obecnej. W 1975 r. obronił pracę doktorską pt. „Estymacja ogólnych modeli liniowych i autoregresyjnych". W 1986 r.

obronił rozprawę habilitacyjną pt. „Liniowe modele statystyczne" i otrzymał stopień doktora habilitowanego nauk ekonomicznych. W 1995 r. otrzymał tytuł naukowy profesora oraz stanowisko profesora nadzwyczajnego Uniwersytetu Łódzkiego. W 1998 r. uzyskał stanowisko profesora zwyczajnego, na którym pracuje do dziś.

Jest autorem, współautorem oraz redaktorem 14 książek opublikowanych m.in. w wydawnictwach naukowych: PWN, PWE oraz Wydawnictwie Uniwersytetu Łódzkiego. Najważniejsze z nich to: „Nieliniowe modele ekonometryczne", PWN, Warszawa 1990, „Stabilność i wrażliwość modeli i metod ekonometrycznych", Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1995, „Stabilność rynków finansowych a wzrost gospodarczy", Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2002.

Jest również autorem lub współautorem ponad 100 artykułów naukowych, rozdziałów monografii, recenzji i not bibliograficznych, opublikowanych m.in. w Palgrave Macmillan, „Publications Econometriques", „Ekonomiście", „Przeglądzie Statystycznym", „Folia Oeconomica" oraz „Wiadomościach Statystycznych". Przedstawił ponad 100 referatów na konferencjach krajowych i zagranicznych.

Zainteresowania naukowe prof. W. Milo początkowo były skoncentrowane na modelach liniowych i autoregresyjnych, którym poświęcił pracę doktorską. Następnie zajmował się analizą szeregów czasowych, a w książce habilitacyjnej własnościami liniowych modeli statystycznych. Po habilitacji badał stabilność, odporność oraz wrażliwość modeli i metod ekonometrycznych. Analizował własności estymacyjno-predykcyjne, zarówno klasycznych jak i nieklasycznych predyktorów, tworzonych na podstawie liniowych i nieliniowych modeli ekonometrycznych.

Od połowy lat dziewięćdziesiątych widoczne jest poszerzenie badań o wątki aplikacyjne dotyczące modelowania i prognozowania rynków pracy, rynków finansowych oraz ich powiązań ze wzrostem gospodarczym. Do najciekawszych należą m.in. te wyniki, które weryfikują prognostyczne własności modeli opartych na teorii Smitha, Ricardo, Marshalla i Keynesa. Potwierdził empirycznie, że modele te nie są prognostycznie gorsze od współcześnie występujących w teorii ekonomii.

W badaniach uwzględniał procesy transformacji gospodarczej i integracji z Unią Europejską. W swych publikacjach dużo uwagi poświęcił również wyborom optymalnych portfeli inwestycyjnych, prognozowaniu kursów walutowych, stóp procentowych, inflacji oraz wzrostowi gospodarczemu.

W ostatnim okresie rozpoczął analizę fundamentalnych zasad wnioskowania, dotyczących przyczynowości i prawdy, a także istoty losowości, chaosu, entropii i ergodyczności w kontekście analiz zjawisk ekonomicznych i finansowych.

Profesor Władysław Milo jest promotorem 12 prac doktorskich, autorem 14 recenzji prac w przewodach doktorskich, 6 recenzji w przewodach habilitacyjnych i 10 recenzji w postępowaniach o tytuł profesora.

Od 1976 r. kierował tematami badań w problemach resortowych, centralnych, a w latach 90. XX w. projektami KBN oraz był głównym wykonawcą w dwóch projektach ACE (Action for Co-operation in the field of Economics) Komisji Europejskiej w Brukseli.

Przebywał na wielu stażach, stypendiach i studiach zagranicznych, m.in. na:
- stypendium Fulbrighta, studia podyplomowe w Duke University, USA (1971/1972),
- stypendiach British Council, University of Bangor, Liverpool, Cambridge (1983, 1986, 1988),
- stypendium DAAD w University of Hagen, Niemcy, 1985.

W ramach współpracy zagranicznej przebywał w l'Université Paris 1, 1986, Manchester University, Oxford University, 1990, University of Amsterdam, 1993 i w European University Institute, Florence, 1996.

Był lub jest członkiem następujących towarzystw naukowych i organizacji naukowych: American Association for the Advancement of Science, Bernoulli Society, Econometric Society, European Economic Association, Komitetu Statystyki i Ekonometrii PAN, Łódzkiego Towarzystwa Naukowego (był sekretarzem oraz przewodniczącym II Wydziału), New York Academy of Science, Polskiego Towarzystwa Ekonomicznego, Polskiego Towarzystwa Filologicznego oraz Polskiego Towarzystwa Matematycznego.

Był redaktorem w dwóch czasopismach:
- „Przegląd Statystyczny" (w latach 1987–2001),
- „Prace Instytutu Ekonometrii i Statystyki" UŁ.

Aktualnie jest Przewodniczącym Rady Redakcyjnej „Przeglądu Statystycznego".

Jest inicjatorem i redaktorem naczelnym serii wydawniczej „Financial Markets. Principles of Modelling, Forecasting and Decision-Making (FindEcon)".

W czasie studiów był członkiem ZSP. W latach 1966–1981 był członkiem ZNP. Od 1981 r. jest członkiem NSZZ „Solidarność". W latach 1989–1990 był konsultantem Zarządu Regionalnego „Solidarność" w Łodzi. Nie był członkiem żadnej partii politycznej.

Za działalność naukowo-dydaktyczną i organizatorską w środowisku akademickim został wyróżniony m.in.: Krzyżem Kawalerskim Orderu Odrodzenia Polski oraz Medalem Komisji Edukacji Narodowej.

Pozanaukowe zamiłowania Profesora to nauka języków obcych, filozofia, kolekcjonerstwo książek, muzyka poważna, zabytki kultury światowej i polskiej, muzea, podróże, poznawanie różnych kultur. Ponadto interesuje się sportem, w szczególności lekkoatletyką.

Jan Jacek Sztaudynger, Piotr Wdowiński – Uniwersytet Łódzki