

Krzysztof Dmytrów*

Mariusz Doszyń**

Uniwersytet Szczeciński

PROGNOZOWANIE SPRZEDAŻY Z ZASTOSOWANIEM ROZKŁADU GAMMA Z KOREKCJĄ ZE WZGLĘDU NA WAHANIA SEZONOWE

STRESZCZENIE

W artykule zbadano efektywność prognoz sprzedaży dużej liczby wyrobów otrzymanych za pomocą dwóch metod: 1) prognozowania na podstawie mediany rozkładu gamma oraz 2) mediany rozkładu gamma z wahaniami sezonowymi. Dokładność prognoz zbadano za pomocą błędów *ex post*, takich jak błąd średniokwadratowy (*MSE – mean squared error*) wraz z jego dekompozycją oraz współczynnik Theila. Otrzymane wyniki wskazują, że z punktu widzenia obciążoności oraz elastyczności lepsze prognozy otrzymano na podstawie mediany rozkładu gamma z wahaniami sezonowymi.

Słowa kluczowe: prognozowanie sprzedaży w przedsiębiorstwie, prognozowanie na podstawie rozkładu gamma, wahania sezonowe.

Wprowadzenie

W gospodarowaniu zapasami w przedsiębiorstwie jedną z najważniejszych rzeczy jest prawidłowa ocena wielkości i rozkładu zapotrzebowania. W zależności od

* Adres e-mail: krzysztof.dmytrow@wneiz.pl.

** Adres e-mail: mariusz.doszyn@wneiz.pl.

tego, czy planuje się zapasy dystrybucyjne, czy zapasy produkcyjne, należy stosować metody prognozowania popytu niezależnego bądź popytu zależnego [Sarjusz-Wolski, 2000]. Popyt zależny wynika z planów produkcyjnych i najczęściej jest prognozowany za pomocą różnych systemów planowania potrzeb materiałowych i podobnych (MRP, MRP II, ERP, *etc.*). Z kolei popyt niezależny wynika z zapotrzebowania zgłaszanego przez odbiorców i ustala się go za pomocą metod prognozowania sprzedaży¹. Można je podzielić na trzy duże grupy:

- prognozowanie na podstawie metod analizy szeregów czasowych,
- prognozowanie z wykorzystaniem ekonometrycznych modeli przyczynowo-opisowych,
- prognozowanie z użyciem rozkładów.

O ile w przypadku pojedynczych produktów zasadne jest podejście indywidualne, o tyle dla przedsiębiorstw mających w ofercie tysiące produktów konieczne jest zaprojektowanie systemu prognoz, który będzie automatycznie klasyfikował szeregi statystyczne i przyporządkowywał im odpowiednią metodę prognozowania. Zaletą takiego podejścia jest możliwość szybkiego wyznaczenia prognoz wielkości sprzedaży dla wielu produktów, główną zaś jego wadą – fakt stosowania podobnych metod dla wielu produktów w danej klasie, co w indywidualnych przypadkach może pogarszać prognozy.

Do prognozowania najczęściej stosowane są metody oparte na analizie szeregów czasowych (modele trendu, modele wygładzania wykładniczego) oraz modele ekonometryczne. Stosunkowo rzadko prognozuje się na podstawie rozkładów, co w wielu przypadkach jest uzasadnione. Z analizy sprzedaży produktów wynika, że często występują sytuacje, w których obserwuje się nieciągłość sprzedaży (czyli istnieje wiele okresów, w których zapotrzebowanie nie występuje, a po nich są okresy występowania sprzedaży). Kształtowanie się sprzedaży w czasie często nie wykazuje istotnych prawidłowości, ponieważ nie można zazwyczaj znaleźć na tyle dobrych zmiennych objaśniających, aby zbudować model ekonometryczny o pożądanych własnościach prognostycznych. W takich sytuacjach w sukurs przychodzi analiza rozkładów wielkości sprzedaży produktów. Prognozy można sporządzać na podstawie pewnych parametrów rozkładów, zarówno empirycznych, jak i teoretycznych.

¹ Przegląd zastosowań wybranych metod prognozowania zawierają m.in. następujące prace [Batóg, Foryś, 2009; Gnat, 2008; Theil, 1961].

W artykule do budowy prognoz wielkości sprzedaży wykorzystano rozkład gamma. Powodem takiego wyboru jest fakt, że jest to najczęściej występujący rozkład zapotrzebowania w gospodarce zapasami [Całczyński, 2000]. Nawet w sytuacjach, gdy rozkład zapotrzebowania nie jest znany, to rozkład gamma jest wskazywany ze względu na swoją uniwersalność. Za jego pomocą można opisać zarówno rozkłady zbliżone do symetrycznych, jak i najczęściej występujące rozkłady prawostronnie asymetryczne. Ponadto wiele rozkładów to szczególne przypadki rozkładu gamma [Tadikamalla, 1984].

1. Opis wykorzystanych danych i procedur

W przeprowadzonym badaniu porównano efektywność prognoz *ex post* sprzedaży na dwanaście miesięcy wyznaczonych na podstawie mediany rozkładu gamma z tymi samymi prognozami, skorygowanymi dodatkowo w każdym miesiącu o wahania sezonowe. Wykorzystane dane statystyczne odnoszą się do przedsiębiorstwa, w którego ofercie znajduje się ponad 12 tys. produktów. Zbudowana baza danych zawiera informacje o miesięcznych wielkościach sprzedaży za okres od maja 2009 roku do sierpnia 2013 roku (52 miesiące). Należy zaznaczyć, że długość szeregów czasowych dla poszczególnych produktów często jest różna. Niektóre produkty były w ofercie przez cały analizowany okres, a niektóre wprowadzono później. Badano tylko te produkty, które znajdują się obecnie w ofercie przedsiębiorstwa. Ostatnie 12 miesięcy (od września 2012 roku do sierpnia 2013 roku) były okresem empirycznej weryfikacji prognoz. Spośród produktów wybrano jedynie te, które były w ofercie przynajmniej przez 24 miesiące, do sierpnia 2012 roku włącznie (czyli przynajmniej od września 2010 roku) oraz te, dla których miesięczna częstość sprzedaży (udział miesięcy z dodatnią wielkością sprzedaży we wszystkich miesiącach badanego okresu) była równa 1. Produkty były więc sprzedawane w każdym miesiącu od pojawienia się sprzedaży. Dla mniejszej częstości sprzedaży rozkłady charakteryzowały się silną asymetrią prawostronną, a każda dodatnia wartość znajdowała się praktycznie w ogniu rozkładu.

Do danych wybranych w przedstawiony sposób dopasowano dwuparametrowy rozkład gamma opisany następującą funkcją gęstości [Krysicki, Bartos, Dyczka, Królikowska, Wasilewski, 2000]:

$$f(x; \lambda, \eta) = \frac{\lambda^\eta}{\Gamma(\eta)} x^{\eta-1} e^{-\lambda x} \quad (1)$$

gdzie $x \geq 0$, $\lambda \geq 0$, $\eta \geq 0$.

Parametry kształtu (λ) oraz skali (η) wyliczono następująco:

$$\lambda = \frac{\bar{x}^2}{S^2(x)} \quad (2)$$

$$\eta = \frac{S^2(x)}{\bar{x}} \quad (3)$$

Zgodność rozkładu sprzedaży z rozkładem gamma weryfikowano na podstawie nieparametrycznego testu Locke'a, w którym przyjmuje się, że parametry rozkładu nie są znane. Test ten odwołuje się do następującej zasady: jeżeli zmienne X i Y są niezależne to zmienne X/Y i $X + Y$ są niezależne wtedy i tylko wtedy, gdy mają rozkład gamma [Locke, 1976; Shapiro, Chen, 2001]. W teście tym próba dzielona jest losowo na dwie podpróby o takiej samej liczebności n . W przypadku nieparzystej liczby obserwacji jedna z nich jest losowo pomijana, co może przyczyniać się do tego, że kolejne zastosowania tego testu mogą prowadzić do nieznacznie różniących się wyników. W kolejnym etapie tworzone są pary następujących zmiennych: $U_i = X_{2i-1} + X_{2i}$; $V_i = \max\left\{\frac{X_{2i-1}}{X_{2i}}, \frac{X_{2i}}{X_{2i-1}}\right\}$. Niezależność zmiennych U_i i V_i bada się na podstawie współczynnika korelacji rang Kendalla [Locke, 1976].

Prognozowanie na poziomie mediany rozkładu gamma ma tę wadę, że dla danego szeregu czasowego wartość zmiennej jest taka sama w każdym prognozowanym okresie. Mimo że tak wyznaczone prognozy są na ogół nieobciążone, to są nieelastyczne. Postanowiono zatem porównać prognozy wyznaczone na podstawie mediany rozkładu gamma z tą samą medianą skorygowaną w każdym miesiącu o wskaźnik sezonowości wyznaczony dla dwóch ostatnich lat (wrzesień 2010 roku – sierpień 2012 roku), na podstawie następującego wzoru:

$$W_i = \frac{\bar{x}_i}{M_s}, i = 1, 2, \dots, 12 \quad (4)$$

gdzie

\bar{x}_i – średnia wielkość sprzedaży w i -tym miesiącu w okresie wrzesień 2010 roku–sierpień 2012 roku,

M_s – mediana sprzedaży w okresie wrzesień 2010 roku–sierpień 2012 roku.

Wyliczone za pomocą wzoru (4) wskaźniki oczyszczono współczynnikiem korygującym, tak ażeby ich suma wyniosła 12. Współczynnik korygujący k obliczono według następującego wzoru:

$$k = \frac{12}{\sum_{i=1}^{12} W_i} \quad (5)$$

Oczyszczone wskaźniki sezonowości oblicza się za pomocą następującego wzoru:

$$W_i^o = W_i k \quad (6)$$

Następnie prognozę wyznaczoną na podstawie mediany rozkładu gamma (M) koryguje się o oczyszczony wskaźnik sezonowości następująco:

$$\hat{x}_i = W_i^o \cdot M \quad (7)$$

Dokładność prognoz zbadano za pomocą współczynnika Theila U oraz dekompozycji błędu średniokwadratowego prognoz (MSE). Współczynnik Theila liczy się następująco [Theil, 1966]:

$$U = \sqrt{\frac{\sum_{T=n+1}^{n+k-1} \left(\frac{y_{P,T+1} - y_{T+1}}{y_T} \right)^2}{\sum_{T=n+1}^{n+k-1} \left(\frac{y_{T+1} - y_T}{y_T} \right)^2}} \quad (8)$$

gdzie:

T – numer okresu prognozowanego,

k – liczba prognozowanych okresów,

n – długość przedziału czasowego próby,

$y_{P,T+1}$ – prognoza w okresie $T + 1$,
 y_T, y_{T+1} – wartości empiryczne zmiennej prognozowanej w okresach T oraz $T + 1$.

Współczynnik U ma wartości dodatnie. Jeżeli jego wartość przekroczy 1, to lepsze prognozy od analizowanego modelu dawałby model naiwny (taki, że $y_{T+1} = y_T$).

Średniokwadratowy błąd prognoz liczy się następująco:

$$MSE = \frac{\sum_{T=n+1}^{n+k} (y_T - y_{PT})^2}{k} \quad (9)$$

Błąd ten ulega dekompozycji na trzy współczynniki [Theil, 1966]:

$$MSE = \underbrace{(\bar{y}_{P,T} - \bar{y}_T)^2}_{\text{obciążoność}} + \underbrace{(s_P - rs_T)^2}_{\text{regresja}} + \underbrace{(1 - r^2)s_T^2}_{\text{zakłócenia}} \quad (10)$$

gdzie:

$\bar{y}_{P,T}$ – wartość średnia prognoz,
 \bar{y}_T – wartość średnia wielkości empirycznych zmiennej prognozowanej,
 s_P – odchylenie standardowe prognoz,
 s_T – odchylenie standardowe wielkości empirycznych zmiennej prognozowanej,
 r – współczynnik korelacji między prognozami a wielkościami empirycznymi zmiennej prognozowanej.

Dzieląc obie strony równania (10) przez MSE , otrzymamy udział błędów wynikających z obciążenia (U_M), regresji (U_R) oraz zakłóceń (U_D) w średniokwadratowym błędzie prognoz:

$$\frac{(\bar{y}_{P,T} - \bar{y}_T)^2}{MSE} + \frac{(s_P - rs_T)^2}{MSE} + \frac{(1 - r^2)s_T^2}{MSE} = 1 \quad (11)$$

Ze względu na specyfikę prognoz wyznaczonych za pomocą rozkładu gamma (bez korekty o wahania sezonowe) nie można policzyć odchylenia standardowego prognoz (s_P) oraz współczynnika korelacji między prognozami i wielkościami empirycznymi zmiennej prognozowanej (r) (wariancja prognoz jest równa zero).

Oznacza to, że nie można obliczyć udziałów regresji (U_R) oraz zakłóceń (U_D) w błędzie MSE . Z tego powodu w przypadku dekompozycji, dla każdej z metod wyznaczono jedynie obciążenie prognoz.

2. Wyniki prognozowania rozkładów

W przykładzie empirycznym wstępnej analizie poddano miesięczną sprzedaż 12076 produktów w okresie od maja 2009 roku do sierpnia 2013 roku. Prognozy budowano na podstawie rozkładu gamma oraz rozkładu gamma z korekcją ze względu na sezonowość, z wykorzystaniem danych dla pierwszych 40 miesięcy, a następnie wyznaczano prognozy *ex post* na kolejne 12 miesięcy. Jak wspomniano, długości szeregów czasowych mogą być różne (krótsze), jeśli produkt był wprowadzany do oferty w późniejszym okresie niż maj 2009 roku.

Prognozy budowano tylko dla tych produktów, dla których nie było podstaw do odrzucenia hipotezy o zgodności rozkładu zmiennej z rozkładem gamma. Zastosowano nieparametryczny test Locke'a (poziom istotności $\alpha = 0,1$). Jednocześnie brano pod uwagę tylko produkty o częstości sprzedaży równej 1, co oznacza, że od momentu pojawienia się produktu w ofercie sprzedano go w każdym miesiącu. Minimalna przyjęta liczba obserwacji to 24 miesiące, czyli popyt na poszczególne produkty był dodatni przez co najmniej ostatnie dwa lata. Wskaźniki sezonowości wyznaczano na podstawie danych z ostatnich 24 miesięcy jako średnią sprzedaż w danym miesiącu podzieloną przez medianę. Po uwzględnieniu tych kryteriów do analizy przyjęto 448 produktów.

Do porównania efektywności prognoz *ex post* posłużono się średniokwadratowym błędem absolutnym MSE oraz współczynnikiem Theila (U). Dokonano również dekompozycji błędu MSE , porównując obciążenie każdej z metod oraz analizując udziały pozostałych rodzajów błędów jedynie dla prognoz wyznaczonych na podstawie mediany z korekcją ze względu na sezonowość. Jak powiedziano, ze względu na zerową wariancję prognoz wyznaczanych na podstawie mediany (bez korekcji o wahania sezonowe), dokonanie pełnej dekompozycji błędów prognoz dla tej metody nie jest możliwe. Kształtowanie się rozkładów błędów *ex post* wyznaczone na podstawie każdej z metod przedstawiono na rysunkach 1–5, natomiast statystyki pozycyjne błędów zestawiono w tabeli 1. Ze względu na silną asymetrię rozkładów analizowano tylko miary pozycyjne. Kwartyłe liczono na podstawie szeregów szczegółowych.

Tabela 1. Statystyki opisowe błędów charakteryzujących efektywność prognoz *ex post* wyznaczonych na podstawie rozkładu gamma oraz rozkładu gamma z korekcją ze względu na sezonowość

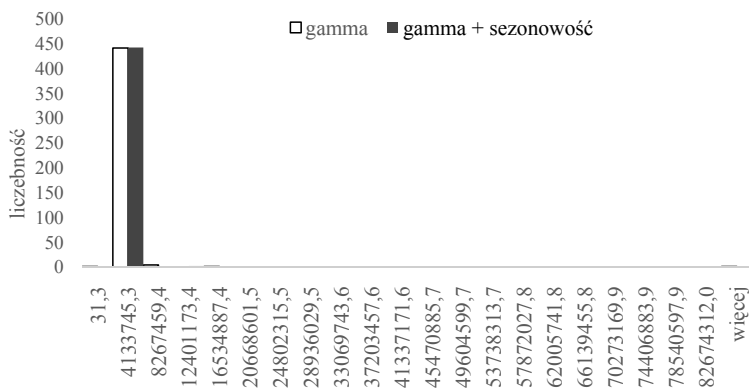
Rodzaj modelu	Rodzaj błędu	MSE	U	U_M	U_R	U_D
Gamma	$Q_{1,4}$	510,168	0,615	0,003	-	-
	M	1301,295	0,790	0,014		
	$Q_{3,4}$	4385,929	1,017	0,033		
	Q	1937,880	0,201	0,015		
	V_Q	1,489	0,254	1,046		
	A_2	0,592	0,129	0,253		
Gamma z sezonowością	$Q_{1,4}$	659,664	0,274	0,002	0,008	0,034
	M	1756,806	0,549	0,010	0,022	0,049
	$Q_{3,4}$	5492,337	1,247	0,026	0,034	0,066
	Q	2416,337	0,487	0,012	0,013	0,016
	V_Q	1,375	0,887	1,179	0,605	0,317
	A_2	0,546	0,435	0,341	-0,041	0,020

- brak możliwości obliczenia.

Źródło: opracowanie własne.

Pierwszy kwartył, mediana i trzeci kwartył rozkładu błędu MSE miały niższe wartości w prognozach wyznaczonych na podstawie rozkładu gamma. Można zatem stwierdzić, że z punktu widzenia błędu MSE nieco korzystniejsze prognozy daje mediana rozkładu gamma (bez korekcji ze względu na wahania sezonowe). Pozytywny współczynnik zmienności oraz asymetria rozkładu były natomiast niższe dla rozkładu błędu MSE odpowiadającego prognozom z wykorzystaniem rozkładu gamma z korekcją ze względu na wahania sezonowe, przy czym różnice były tutaj niewielkie. Rozkłady błędu MSE w każdym przypadku cechowały się silną asymetrią prawostronną z powodu występowania wartości ekstremalnych, czyli dużych wartości błędu prognoz dla pojedynczych produktów (zob. rysunek 1).

Rysunek 1. Rozkład błędów MSE wyznaczonego na podstawie rozkładu gamma oraz rozkładu gamma z korekcją ze względu na sezonowość

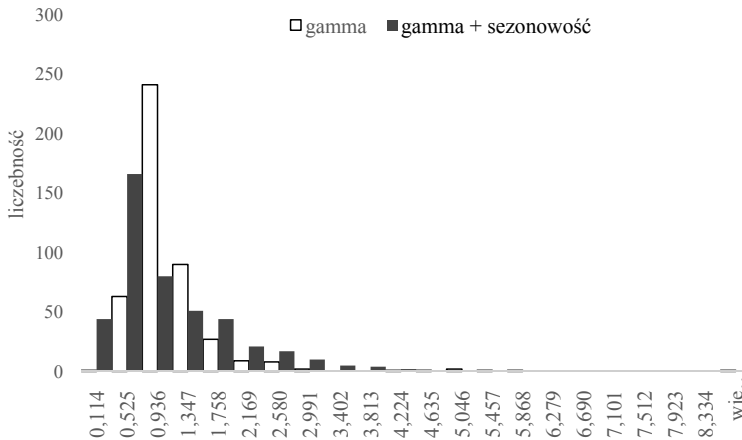


Źródło: opracowanie własne.

Za podstawowe kryterium do porównywania efektywności prognoz *ex post* przyjęto pierwiastek ze współczynnika Theila (U). Jest to inna niż zazwyczaj stosowana wersja tego współczynnika, przedstawiona w pracy [Theil, 1966]. Współczynnik ten zastosowano dlatego, że jego własności są korzystniejsze niż właściwości najczęściej używanego współczynnika, opisanego w pracy [Theil, 1961].

Parametry empirycznego rozkładu współczynnika Theila U nie prowadzą do jednoznacznych wniosków co do efektywności porównywanych metod. Pierwszy kwartył i mediana były niższe w prognozach obliczanych na podstawie rozkładu gamma z uwzględnieniem wahań sezonowych. Trzeci kwartył był natomiast niższy w prognozach sporządzonych na podstawie rozkładu gamma. Zmienność błędów *ex post* oraz asymetria rozkładu błędów były niższe również wtedy, gdy prognozy nie były korygowane o wskaźniki sezonowości. Świadczą o tym pozycyjne miary przedstawione w tabeli 1. Wykresy rozkładów błędów porównano na rysunku 2.

Rysunek 2. Rozkład współczynnika Theila (U) wyznaczonego na podstawie rozkładu gamma oraz rozkładu gamma z korekcją ze względu na sezonowość



Źródło: opracowanie własne.

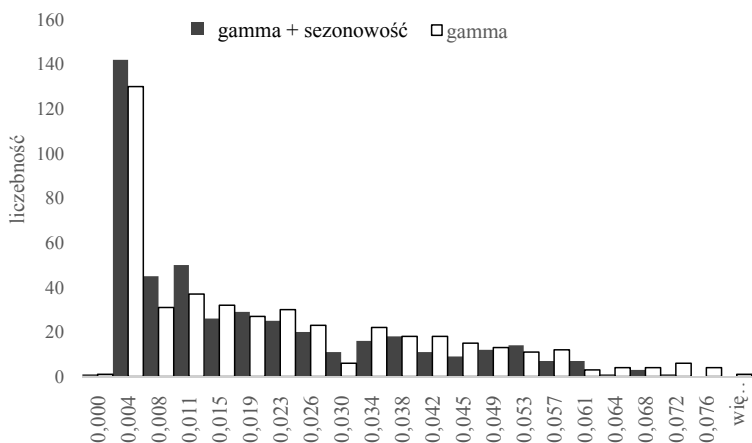
Po dekompozycji błędu MSE porównano udziały błędów wynikających z obciążenia prognoz. Z punktu widzenia tego rodzaju błędów nieco korzystniej kształtują się prognozy obliczone na podstawie rozkładu gamma z uwzględnieniem wahań sezonowych, przy czym różnice te nie są zbyt duże. Wszystkie wyznaczone kwartyle rozkładu błędów są niższe dla prognoz z korekcją ze względu na sezonowość. Zmienność i asymetria rozkładu błędów natomiast mają niższe wartości dla prognoz na podstawie mediany rozkładu gamma. Wnioski te potwierdza rozkład współczynnika obciążenia (rysunek 3).

Jak wspomniano, prognozy wyznaczone na podstawie mediany rozkładu gamma w całym okresie prognozowanym mają stałe wartości, a więc są nieelastyczne (wariancja prognoz jest równa zero). Dla tego typu prognoz nie można wyznaczyć błędów wynikających z „regresji” i „zakłóceń” prognoz (wzór (10)). Tego rodzaju błędy natomiast poddano analizie dla prognoz korygowanych o wskaźniki sezonowości. Rozkłady ich udziałów w MSE przedstawiono na rysunku 4 i 5.

Na podstawie analizy rozkładu współczynnika U_R wyznaczonego na podstawie rozkładu gamma z korekcją ze względu na sezonowość można stwierdzić, że dominują przedziały o niskim poziomie tego wskaźnika, co należy uznać za zjawisko

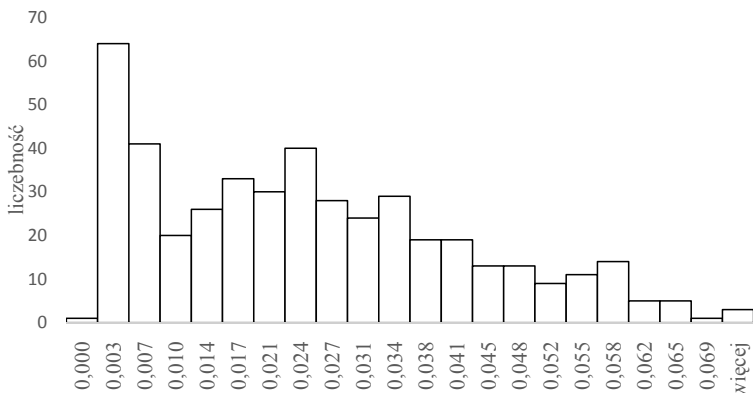
pozytywne. Z kolei rozkład współczynnika U_D świadczy o dość znacznej jednorodności błędów prognoz wynikających z „zakłóceń”.

Rysunek 3. Rozkład współczynnika obciążenia (U_M) wyznaczonego na podstawie rozkładu gamma oraz rozkładu gamma z korekcją ze względu na sezonowość



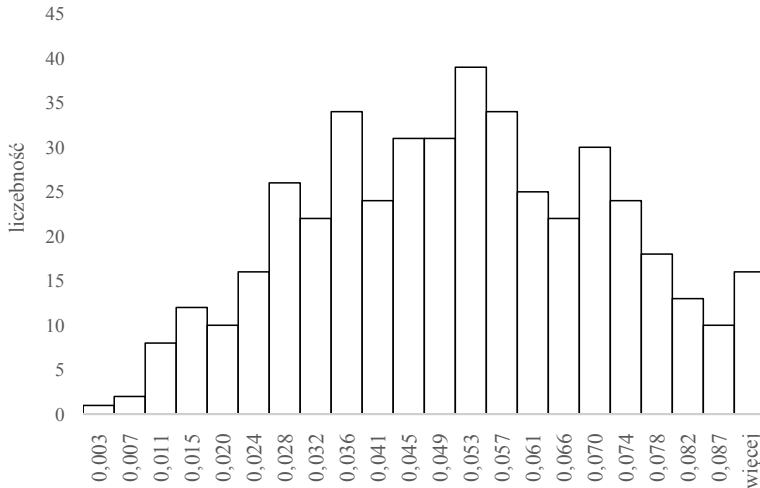
Źródło: opracowanie własne.

Rysunek 4. Rozkład współczynnika regresji (U_R) wyznaczonego na podstawie rozkładu gamma z korekcją ze względu na sezonowość



Źródło: opracowanie własne.

Rysunek 5. Rozkład współczynnika zakłóceń (U_D) wyznaczonego na podstawie rozkładu gamma z korekcją ze względu na sezonowość



Źródło: opracowanie własne.

Podsumowanie

Porównanie efektywności prognoz *ex post* wyznaczonych na podstawie mediany rozkładu gamma oraz mediany rozkładu gamma korygowanej o wskaźniki sezonowości nie prowadzi do jednoznacznych wniosków. Z punktu widzenia błędu średniokwadratowego (*MSE*) nieco lepsze prognozy wyznaczono na poziomie mediany rozkładu gamma. Z kolei współczynnik Theila nieznacznie „faworyzuje” metodę z korekcją ze względu na wahania sezonowe, szczególnie z punktu widzenia obciążenia prognoz. Na korzyść tej metody przemawia również ich większa elastyczność i zazwyczaj mniejsza niezgodność kierunków prognoz i wartości empirycznych. Reasumując, efektywność każdej z opisywanych metod zależy od właściwości zbioru danych (własności szeregów czasowych) i w każdym przypadku powinna być analizowana na przykład za pomocą stosowanych w niniejszym artykule miar efektywności prognoz *ex post*. Należy także zauważyć, że przedstawione metody można wykorzystać do prognozowania nie tylko zmiennych charakteryzujących sprzedaż.

Literatura

- Batóg B., Foryś I. (2009), *Prognozowanie zużycia ciepłej i zimnej wody w spółdzielczych zasobach mieszkaniowych*, w: *Metody ilościowe w ekonomii*, Studia i Prace WNEiZ nr 2, Szczecin.
- Całczyński A. (red.) (2000), *Elementy badań operacyjnych w zarządzaniu*, t. 1, Radom.
- Gnat S. (2008), *Prognozowanie dochodów ze sprzedaży tygodników lokalnych – wybrane podejścia*, w: *Metody ilościowe w ekonomii*, Studia i Prace WNEiZ nr 2, Szczecin.
- Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. (2000), *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. I. Rachunek prawdopodobieństwa*, Wydawnictwo Naukowe PWN, Warszawa.
- Locke C. (1976), *A Test for the Composite Hypothesis that Population Has a Gamma Distribution*, *Communications in Statistics* 4.
- Sarjusz-Wolski Z. (2000), *Sterowanie zapasami w przedsiębiorstwie*, PWE, Warszawa.
- Shapiro S.S., Chen L. (2001), *Composite Test for the Gamma Distribution*, „*Journal of Quality Technology*” Vol. 33, No. 1.
- Tadikamalla P.R. (1984), *A Comparison of Several Approximations to the Lead Time Demand Distribution*, „*International Journal of Management Science*”, Vol. 12, No. 6.
- Theil H. (1966), *Applied Economic Forecasting*, North-Holland, Amsterdam.
- Theil H. (1961), *Economic Forecasting and Policy*, North-Holland, Amsterdam.

SALES FORECASTING BASED ON GAMMA DISTRIBUTION WITH SEASONAL ADJUSTMENTS

Abstract

In the article efficiency of forecasts obtained by means of two types of method was analysed. Forecasts were based on both median of gamma distribution and median of gamma distribution with seasonal adjustments. To evaluate forecasts' effectiveness such *ex post* errors of forecasts as *MSE* (mean squared error) and Theil coefficient were applied. Decomposition of *MSE* was also introduced. According to empirical results, with respect to unbiasedness and elasticity, forecasts based on median of gamma distribution with seasonal adjustments seem to be more efficient.

Translated by Mariusz Doszyń

Keywords: sales forecasting, forecasting based on gamma distribution, seasonal adjustments.

Kod JEL: C53, C12, C13.

