

Katarzyna CHUDY-LASKOWSKA, Maria WIERZBIŃSKA

Prognozowanie transportu towarowego

Transport stał się jednym z ważniejszych działów gospodarki. Polskie firmy transportowe dokonują przemieszczeń towarów zarówno wewnątrz kraju, jak i międzynarodowych. Liczy się zwłaszcza tranzyt w kierunkach południe-północ oraz wschód-zachód i odwrotnie¹.

Celem artykułu było przygotowanie prognozy dla transportu towarowego (przewozu ładunków) w Polsce od września 2010 r. do marca 2011 r.

CHARAKTERYSTYKA DANYCH STATYSTYCZNYCH I IDENTYFIKACJA ICH SZEREGU CZASOWEGO

Do analiz wykorzystano dane dotyczące przewozu ładunków w Polsce od lipca 2001 r. do lipca 2010 r.²

Analizując dane dotyczące przewozu ładunków zaobserwować można niewielką tendencję wzrostową oraz wahania sezonowe i przypadkowe. Począwszy od 2001 r. wzrastała wielkość przewozu ładunków. Najwyższe wartości odnoto-

¹ www.polandtransport.eu

² www.stat.gov.pl

wywano najczęściej we wrześniu (czasem w październiku), natomiast naniższe — zazwyczaj w styczniu i lutym. Z przeprowadzonego testu ANOVA³ wynika, że istnieje statystyczna różnica w średnich przewozach ładunków w poszczególnych miesiącach, o czym świadczy wartość $p < \alpha$, ($F=2,91$; $p=0,005621$, $\alpha = 0,05$).

Do wyodrębnienia podstawowych składowych szeregu wykorzystano metodę dekompozycji sezonowej (Makridakis, Wheelwright, Mc Gee, 1983). Ogólna idea dekompozycji sezonowej jest prosta. Celem metody jest wyodrębnienie składników szeregu i ich dekompozycja na efekt trendu, efekty sezonowe i pozostałą zmienność.

O przewozie ładunków można powiedzieć, że składa się z czterech składników: sezonowego (określanego jako S_t , gdzie t oznacza określony punkt w czasie); trendu (T_t); cyklicznego (C_t); losowego błędu lub nieregularności (I_t). Różnica między składnikiem cyklicznym a sezonowym polega na tym, że ten drugi

³ Analiza wariancji (ANOVA) jest zbiorem metod umożliwiających porównanie średnich w kilku badanych grupach i sprawdzenie, czy założony (wybrany) czynnik ma wpływ na średni poziom analizowanej zmiennej. Do badań wykorzystano test F.

pojawia się w regularnych (sezonowych) odstępach, podczas gdy czynniki cykliczne mają zwykle dłuższy, niekiedy zmieniający się okres. W metodzie Census I⁴ trend i wahania cykliczne zostają zwyczajowo połączone w składnik wahań długookresowych i trendu (trend-cykl, TC_I). Szczegółowa zależność funkcyjna między tymi składnikami może przybierać różne formy (addytywną lub multiplikatywną bądź ich kombinację).

Dekompozycja składowych badanego szeregu pozwala sformułować hipotezę, że przewóz ładunków charakteryzuje się tendencją wzrostową o charakterze wykładniczym. Rozkład zaś składnika losowego (białego szumu) jest rozkładem normalnym, $p > \alpha$ ($p = 0,33852$), a wahania okresowe mają charakter addytywny. Ostateczny zmodyfikowany szereg nieregularny powstał przez zastąpienie wartości w ostatecznym składniku nieregularnym zerem (model addytywny) w przypadku, gdy były one zidentyfikowane jako ekstremalne. Ekstrema w składniku sezonowym uwidaczniają się dzięki temu, że od trendu i wahań długookresowych odejmuje się wartość szeregu pierwotnego, skorygowanego ze względu na liczbę dni handlowych i zmienność wstępną. Po czym na tak ukształtowanym szeregu przeprowadza się analizę wariancji względem miesiąca w celu przetestowania obecności stabilnej i istotnej sezonowości.

⁴ Statistica, Volume III: Statistics II, StatSoft, Inc. Tulsa OK. (1995), s. 3391—3400.

METODY BADAWCZE

Do prognozowania przewozu ładunków wykorzystano dwie metody: wygładzanie wykładnicze oraz metodologię ARIMA.

Prognozy otrzymane za pomocą wygładzania wykładniczego

Istota wygładzania wykładniczego polega na tym, że szereg czasowy zmiennej prognozowanej wygładza się za pomocą średniej ruchomej ważonej, a wagi ustalone są według funkcji wykładniczej. Prognoza oparta jest na średniej ważonej aktualnych i retrospekcyjnych wartościach szeregu (Aczel, 2000). Wygładzanie wykładnicze może być oparte na różnych modelach, odpowiednich do rodzaju składowych analizowanego szeregu czasowego.

Do określenia opóźnień sezonowych w modelu posłużył korelogram, czyli wykres funkcji autokorelacji szeregu. Z korelogramu wynika, że opóźnienie sezonowe wynosi dwanaście, czyli dla co dwunastej obserwacji notuje się wartości maksymalne szeregu.

W kolejnym etapie został oszacowany model wygładzania wykładniczego z trendem wykładniczym i wahaniami addytywnymi oraz multiplikatywnymi. Dekompozycja sezonowa wskazuje, że wahania są addytywne, lecz w ostatnich dwóch latach amplituda wahań wyraźnie zwiększa się, nie biorąc pod uwagę trendu. Dlatego zostaną oszacowane dwa modele i ten, który będzie obarczony mniejszymi błędami⁵, posłuży za model prognostyczny.

W przypadku modeli wygładzania wykładniczego kluczową sprawą jest wybór wartości parametrów modelu. W literaturze przedmiotu podaje się kilka sposobów wyboru parametrów (Gardner, 1985). Jeśli poszczególne składowe zmieniają się szybko, to uważa się, że wartości parametrów wygładzania należy ustalić na poziomie bliskim jedności, w przeciwnym wypadku na poziomie bliskim zeru. Wartości parametrów można dobrać także metodą eksperymentu, minimalizując wybrane błędy.

TABL. 1. ZESTAWIENIE BŁĘDÓW DLA WYBRANYCH MODELI WYGŁADZANIA WYKŁADNICZEGO

Rodzaje błędów	Modele	
	trend wykładniczy wahania addytywne parametry: $\alpha=0,7; \delta=0,1; \gamma=0,1$	trend wykładniczy wahania multiplikatywne parametry: $\alpha=0,7; \delta=0,1; \gamma=0,1$
Błąd średni	-0,012	-0,019
Średni błąd bezwzględny	0,894	0,875
Suma kwadratów	185,405	172,341
Średni błąd kwadratowy	1,685	1,567
Średni błąd procentowy	-0,14073	-0,14088
Średni bezwzględny błąd procentowy	3,119	3,043

α — parametr charakteryzujący stopień wygładzania jest niezbędny we wszystkich modelach; δ — parametr wygładzania sezonowego i określany tylko w przypadku wyboru modeli sezonowych; γ — parametr wygładzania trendu liniowego i wykładniczego; φ — parametr określany w przypadku trendu gasnącego.

Źródło: opracowanie własne.

Analiza błędów wskazuje, że bardziej dopasowany jest model trendu wykładniczego z wahaniami multiplikatywnymi. Cztery z przedstawionych błędów przyjmują mniejszą wartość. Tak więc do prognozowania przewozu ładunków od września 2010 r. do marca 2011 r. posłuży model z trendem wykładniczym oraz wahaniami multiplikatywnymi. Postać modelu przedstawia się następująco:

$$F_{t-1} = \alpha(y_{t-1} - C_{t-1-r}) + (1 - \alpha)(F_{t-2} - S_{t-2}) \quad (1)$$

$$S_{t-1} = \gamma(F_{t-1} - F_{t-2}) + (1 - \gamma)S_{t-2} \quad (2)$$

⁵ Zestaw błędów, spośród których wybrany został najbardziej precyzyjny model, przedstawia tabl. 1.

$$C_{t-1} = \delta \frac{y_{t-1}}{F_{t-1}} + (1 - \gamma) C_{t-1-r} \quad (3)$$

gdzie:

- F_{t-1} — odpowiednik wygładzonej wartości otrzymanej z prostego modelu wygładzania wykładniczego (ocena wartości średniej),
- S_{t-1} — ocena przyrostu trendu w momencie lub okresie $t-1$,
- C_{t-1} — ocena wskaźnika sezonowości w momencie lub okresie $t-1$,
- r — długość cyklu sezonowego — liczba faz,
- α, γ, δ — parametry modelu przyjmujące wartości z przedziału od [0 do 1].

Równanie prognozy w momencie lub okresie $t > n$ (n jest liczbą wyrazów szeregu zmiennej prognozowanej) dla wersji multiplikatywnej modelu dane wzorem:

$$y_t^* = [F_n + S_n(t - n)] C_{t-r} \quad (4)$$

jest wyznaczoną na podstawie szeregu czasowego średnią ilorazów odpowiadających tej samej fazie cyklu sezonowego wartości zmiennej prognozowanej i wygładzonych wartości trendu (Cieślak, 2001).

Reszty modelu charakteryzują się rozkładem normalnym $p < \alpha$, ($p = 0,10961$), a więc model jest poprawnie oszacowany.

**TABL. 2. PROGNOZA PRZEWOZU ŁADUNKÓW
OD WRZEŚNIA 2010 R. DO MARCA 2011 R.**

Miesiące	Prognoza
Wrzesień 2010	36,2
Październik 2010	37,5
Listopad 2010	36,1
Grudzień 2010	32,9
Styczeń 2011	31,0
Luty 2011	30,7
Marzec 2011	34,3

Źródło: jak przy tabl. 1.

Wyniki otrzymanych prognoz zachowują tendencję oraz wahania szeregu obserwowanego.

Prognozowanie na podstawie modelu ARIMA

W identyfikacji zjawisk gospodarczych można wyróżnić czynniki, które wskazują na występowanie opóźnień w przebiegu niektórych zjawisk. Zasada echa głosi, że „popyt restytucyjny na trwałe dobra konsumpcyjne jest echem popytu z lat wcześniejszych, przy czym odbicie następuje w okresach równych okresowi zużycia tych dóbr” (Dittmann, 2005). Biorąc pod uwagę dane dotyczące przewozu ładunków zauważyć można wyraźne wahania okresowe, które pojawiają się w odstępach dwunastomiesięcznych, gdzie najwyższe wartości przewozu ładunków odnotowuje się najczęściej w październiku, a najniższe — zazwyczaj w lutym. W tego rodzaju sytuacjach znajdują zastosowanie modele autoregresyjne, w których wartość zmiennej prognozowanej jest funkcją wartości tej zmiennej w momentach lub okresach poprzedzających okres badany oraz składnika losowego.

Model ARIMA (Box, Jenkins, 1983) został wprowadzony w 1976 r. Może być stosowany do modelowania szeregów stacjonarnych, w których występują jedynie wahania losowe wokół średniej lub niestacjonarnych, sprowadzanych do stacjonarnych. Budowa modelu oparta jest na zjawisku autokorelacji. Model ARIMA zawiera dwa podstawowe procesy: autoregresji i średniej ruchomej. W niektórych okolicznościach występują one razem. Postać modelu autoregresji jest następująca:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t \quad (5)$$

Równanie (5) nazywa się procesem autoregresji rzędu p , AR(p),
gdzie:

Y_t, Y_{t-1}, Y_{t-p} — wartość zmiennej objaśnianej w okresie $t, t-1, t-2, t-p$,
 $\varphi_0, \varphi_1, \varphi_2, \varphi_p$ — parametry modelu,
 e_t — reszta modelu dla okresu t ,
 p — wielkość opóźnienia.

Szeregi czasowe składają się zazwyczaj z obserwacji wzajemnie zależnych tak, że można oszacować współczynniki modelu, które opisują kolejne elementy szeregu na podstawie opóźnionych w czasie poprzednich elementów. Każda obserwacja jest sumą składnika losowego oraz kombinacji liniowej poprzednich obserwacji. Proces AR będzie stabilny, jeśli parametry będą należeć do przedziału $(-1, 1)$. Jest to tzw. wymóg stacjonarności szeregu.

Postać modelu średniej ruchomej (MA):

$$Y_t = \theta_0 + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (6)$$

Równanie (6) nazywa się procesem średniej ruchomej⁶ rzędu q , MA(q),
gdzie:

Y_t — wartość zmiennej objaśnianej w okresie t ,
 $\theta_0, \theta_1, \theta_2, \theta_q$ — parametry modelu,
 $e_t, e_{t-1}, e_{t-2}, e_{t-q}$ — reszty modelu w okresach $t, \dots, t-q$,
 q — wielkość opóźnienia.

W średniej ruchomej każda obserwacja składa się ze składnika losowego oraz kombinacji liniowej składników losowych z przeszłości. Równanie średniej ruchomej da się zapisać w formie autoregresyjnej. Można to wykonać tylko wtedy, gdy parametry średniej ruchomej spełniają określone warunki, tzn. jeśli model jest odwracalny. Do osiągnięcia większej elastyczności w dopasowaniu modelu do szeregu czasowego celowe jest połączenie obu modeli:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (7)$$

Proces taki nazywa się procesem autoregresji i średniej ruchomej rzędu (p i q), co w skrócie zapisuje się ARMA (p, q).

W modelu zakłada się, że wartość zmiennej prognozowanej w momencie lub okresie t zależy od poprzedniej jej wielkości oraz od różnicy pomiędzy przeszłą wartością rzeczywistą zmiennej prognozowanej a jej wartością uzyskaną z modelu. Model ARIMA zawiera parametry autoregresyjne, średniej ruchomej oraz wprowadza operator różnicowania. W modelu ARIMA (p, d, q) wyróżniamy

⁶ Termin „średnia ruchoma” jest nieco mylący, ponieważ suma wag nie jest równa jedności, jednakże jest on ogólnie przyjęty i używany w tym modelu.

więc trzy parametry: autoregresyjne p , rząd różnicowania d , średniej ruchomej q . Wymagane jest, aby wejściowy szereg dla metody ARIMA był stacjonarny, tzn. powinien mieć stałą w czasie średnią, wariancję i brak autokorelacji. Dlatego szereg zazwyczaj potrzebuje różnicowania aż do otrzymania stacjonarności — ile razy szereg powinien być różnicowany wyraża parametr d . Bardzo rzadko liczby parametrów p i q muszą być większe od 2. Są także modele sezonowe, w których dodatkowo określa się trzy parametry sezonowości ARIMA (p_s, d_s, q_s): sezonowe autoregresyjne p_s , sezonowe różnicowania d_s oraz sezonowe średniej ruchomej q_s .

Podstawowy szereg danych o przewozach ładunków charakteryzuje się tendencją wzrostową, wahaniami sezonowymi i przypadkowymi, tak więc nie ma ani stałej średniej, ani wariancji. Dlatego też, aby doprowadzić szereg do stacjonarności, czego wymagają założenia metodologiczne, należało zlogarytmować dane oraz dokonać jednokrotnego różnicowania. Po tych zabiegach szereg można było poddać modelowaniu za pomocą procesu ARIMA.

Po wstępnych przekształceniach szereg uzyskał stacjonarność: ma stałą średnią i wariancję ($p=0,999069$). Najlepszym zestawem parametrów dla modelu

ARIMA po sprawdzeniu kilku kombinacji okazał się: jeden parametr autoregresyjny p , jeden parametr różnicowania (szereg był raz różnicowany) d , jeden parametr średniej ruchomej q oraz jeden parametr autoregresyjny sezonowy p_s , tzn. model ARIMA (1, 1, 1), (1, 0, 0).

TABL. 3. OCENY PARAMETRÓW MODELU

Parametry	Wielkość parametru	Błąd standardowy	Statystyka t (106)	p
$p(1)$	0,62	0,09	6,95	0,000000
$q(1)$	0,95	0,03	31,58	0,000000
$p_s(1)$	0,66	0,10	6,86	0,000000

Źródło: jak przy tabl. 1.

Wszystkie parametry są istotne statystycznie, ponadto sprawdzono autokorelogram, który nie wykazuje żadnych związków pomiędzy obserwacjami (brak autokorelacji), co jest wyznacznikiem poprawności dopasowania modelu.

Dla tak opracowanego modelu wyznaczono prognozę ładunków na kolejne okresy, którą przedstawiono w tabl. 4 i na wykr. 8.

Wyniki otrzymanych prognoz zachowują tendencję oraz wahania szeregu obserwowanego. W okresie prognozowanym zachowują się podobnie jak w prognozie wykonanej modelem wykładniczym.

TABL. 4. PROGNOZA PRZEWOZU ŁADUNKÓW OD WRZEŚNIA 2010 R. DO MARCA 2011 R. OTRZYMANA ZA POMOCĄ MODELU ARIMA (1, 1, 1), (1, 0, 0)

Okresy	Prognoza		
	punktowa	przedziałowa	
		dolna 95%	górna 95%
Wrzesień 2010	37,51	33,82	41,61
Październik 2010	38,03	33,57	43,09
Listopad 2010	37,34	32,66	42,70
Grudzień 2010	34,48	30,02	39,61
Styczeń 2011	29,80	25,87	34,32
Luty 2011	30,86	26,74	35,61
Marzec 2011.....	34,02	29,44	39,31

Źródło: opracowanie własne.

PORÓWNANIE PROGNOZ WYZNACZONYCH DWOMA METODAMI

Prognozowanie transportu towarowego przedstawiono na przykładzie szeregu czasowego danych z Polski, były to dane miesięczne. Analiza kształtowania się danych przyjętych do badań umożliwiła wykorzystanie odpowiednich procedur

badawczych. Obliczono również miary prognoz, które umożliwiły odpowiedź na pytanie, które wyniki prognoz mogą być najefektywniejsze (wykr. 9).

Wyniki prognoz otrzymanych obiema metodami zachowują podobne tendencje. Do tych wyników jednak należy podchodzić z dużą ostrożnością. Powszechnie wiadomo, że prognozowanie jest obciążone błędami natury obiektywnej, jak i subiektywnej. Otrzymane prognozy mogą jednak stanowić źródło informacji dla podejmujących decyzje w obrębie rynku transportu towarowego. Przeprowadzone badania mogą dostarczyć również informacji dotyczących słabych i mocnych stron badanego rynku.

PORÓWNANIE PROGNOZ Z DANYMI RZECZYWISTYMI ZA CZWARTY KWARTAŁ 2010 R.

Można też było sprawdzić, jakimi błędami obarczone są prognozy skonstruowane na podstawie wybranych modeli w stosunku do danych rzeczywistych za IV kwartał 2010 r. Porównanie błędów *ex post* zawiera tabl. 5.

W metodzie ARIMA błędy prognoz były mniejsze w przypadku października i listopada 2010 r., natomiast bardziej trafne prognozy dla września i grudnia 2010 r. dała metoda wykładniczego. Prognozy wyznaczone obiema metodami charakteryzowały się błędami mniejszymi niż 3%, wyjątek stanowi wartość prognozy wyznaczona metodą ARIMA dla grudnia 2010 r., gdzie błąd wynosił 7%.

TABL. 5. PORÓWNANIE BŁĘDU WZGLĘDNEGO PROGNOZ (*ex post*)

Miesiące	Wartość faktyczna	Model prognozy		Błąd względny (%) (<i>ex post</i>)	
		ARIMA	wygładzanie wykładnicze	ARIMA	wygładzanie wykładnicze
Wrzesień	36,50	37,51	36,2	-2,8	0,8
Październik	38,40	38,03	37,5	1,0	2,3
Listopad	36,80	37,34	36,1	-1,5	1,9
Grudzień	32,20	34,48	32,9	-7,1	-2,2

Źródło: opracowanie własne.

dr Katarzyna Chudy-Laskowska, dr Maria Wierzbńska — Politechnika Rzeszowska

LITERATURA

- Aczel A. D. (2000), *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa
- Box G. E. P., Jenkins G. M. (1983), *Analiza szeregów czasowych — prognozowanie i sterowanie*, PWN, Warszawa
- Cieślak M. (2001), *Prognozowanie gospodarcze: metody i zastosowania*, PWN, Warszawa
- Dittman P. (2005), *Prognozowanie w przedsiębiorstwie: metody i ich zastosowania*, Wydawnictwo Oficyna Ekonomiczna, Kraków
- Gardner E. S. Jr. (1985), *Exponential smoothing: The state of the art*. „Journal of Forecasting”, 4, John Wiley & Sons, New York
- Makridakis S., Wheelwright S. C., Mc Gee V. E. (1983), *Forecasting: Methods and Applications*, 2nd edition, John Wiley & Sons, New York

SUMMARY

The aim of the article is to work out a short-term forecast on freight in Poland. The forecast process was conducted on the basis of monthly data and comprises the period from July 2001 to July 2010. The forecast concerns the fourth quarter 2010 and the first quarter 2011. In the researched series there were separated and identified the following elements: trends, seasonal and accidental fluctuations. Exponential smoothing and ARIMA methods were used to forecast the freight. Forecast measures were also calculated to assess which results can be the most likely. The prediction results were compared with the actual CSO data for the fourth quarter 2010.

РЕЗЮМЕ

Целью статьи является представление краткосрочного прогноза транспорта товаров. Обследование было проведено на основе месячных данных с июля 2001 г. по июль 2010 г. для периода сентябрь 2010 г. — март 2011 г. В обследуемом ряду были выделены и идентифицированы следующие составные элементы: тренды, сезонные и случайные колебания.

Для прогнозирования были использованы методы экспоненциального выравнивания и модель АРИМА. Были исчислены также меры прогнозов, что сделало возможным ответ на вопрос, какие результаты прогнозирования могут быть наиболее вероятными. Было сделано сравнение прогнозов за IV квартал 2010 г. с фактическими данными опубликованными в изданиях ЦСУ.