

Prognozowanie popytu na usługi transportu intermodalnego

Demand forecast for intermodal transport services

Celem artykułu było opracowanie modelu prognozowania popytu na usługi transportowe operatora intermodalnego. Na podstawie udostępnionych danych, dotyczących liczby eksportowanych kontenerów, przedstawiono proces opisania zjawiska na podstawie jego przeszłych obserwacji, a także jego ewolucji w przyszłości. Zaproponowano dwa modele: regresji oraz ARIMA. Dla każdego z nich dokonano predykcji przyszłych obserwacji. Otrzymane wartości prognoz porównano i na tej podstawie wybrano model opisujący lepiej badane zjawisko, tzn. dający mniejszy błąd prognozy.

Słowa kluczowe:

transport intermodalny, model ARIMA, model regresji, prognozowanie popytu.

In the article it was presented a model of demand forecast for intermodal operator transport services. Based on the shared data on the number of exported containers is presented the process of describing the observable occurrence on its past observations, as well as its evolution in the future. Two models were proposed: Regression and ARIMA. For each of them, was made a prediction of future observations. The received values for the predictions were compared and a model describing a better tested observable occurrence was chosen, i.e. that gives a smaller forecast error.

Key words:

intermodal transport, ARIMA model, demand forecast, Regression model.

Charakterystyka firmy

Przedmiotem badania jest jeden z oddziałów dużego operatora intermodalnego, którego główna działalność skupia się na świadczeniu usług kontenerowych. Główne przewozy realizowane są w oparciu o własną flotę statków i obejmują usługi importowe, szczególnie w kierunku Chin i Tajwanu, oraz eksportowe, przede wszystkim do Stanów Zjednoczonych i Afryki. Transport lądowy jest oparty na dwu gałęziach: drogowej i kolejowej. Przewozy odbywają się w trzech następujących konfiguracjach:

- od drzwi do portu — dostawa odbywa się od klienta do portu transportem lądowym, następnie drogą morską do docelowego portu skąd towar odbierany jest samodzielnie przez zleceniodawcę;
- od drzwi do drzwi — dostawa odbywa się od klienta do portu transportem lądowym, następnie drogą morską do docelowego portu, skąd towar jest

przewożony do punktu wskazanego przez zleceniodawcę;

- od portu do drzwi — dostawa do portu pozostaje w gestii klienta, natomiast pozostała część trasy realizowana jest przez firmę: drogą morską do docelowego portu oraz lądową do magazynu klienta.

Kontenery są przewożone na lądzie za pomocą pojazdów samochodowych oraz koleją w oparciu o dwa połączenia pomiędzy Gdańskiem i Kętami Wrocławskimi oraz Gdańskiem i Sławkowem. Linie te są zsynchronizowane z wpływieniami kontenerowców do portu. Na przykładzie pierwszej z nich dokonano analizy przewozów kolejowych przez pryzmat liczby eksportowanych kontenerów. Są to dane miesięczne, zgromadzone od stycznia 2012 do marca 2018 roku. Ostatnie pięć obserwacji nie wzięło udziału w modelowaniu. Zostały zachowane jako testowe do weryfikacji prognozy.

Metodyka badania

Pierwszym etapem analizy, umożliwiającym podjęcie właściwych decyzji dotyczących modelowania, jest wizualna inspekcja rzeczywistych danych (Bielińska, 2007; Dittmann, 2000) przedstawionych na rysunku 1.

Na wykresie widać wyraźny trend i pewne wahania wynikające ze zmiany popytu lub losowe (przypadkowe). Zostanie to zweryfikowane w dalszej części badania. Szereg czasowy liczy 69 obserwacji, pozwala to na stosowanie zaawansowanych modeli ekonometrycznych, np. ARIMA (Sokołowski, 2016; Czyżycki, Klóńska, 2011).

Model regresji

W pierwszej kolejności została zidentyfikowana analityczna postać tendencji rozwojowej. Na podstawie wykresu zmiennej można stwierdzić, że ma ona postać liniową lub wielomianu drugiego stopnia. Wykorzystując metodę najmniejszych kwadratów wyznaczono funkcję regresji dla zmiennej ozna-

czającej czas (a dokładnie kolejne miesiące) oraz jej kwadratu. Estymacji dokonano w programie Statistica, a otrzymane wyniki przedstawiono w tabeli 1.

Brak statystycznej istotności zmiennej t , którą wskazuje wartość prawdopodobieństwa testowego p , wyższa od założonego poziomu istotności $\alpha = 0,05$ oznacza, że zmienna nie ma wpływu na kształtowanie tendencji rozwojowej, dlatego usunięto ją z modelu i ponownie dokonano estymacji. Wyniki estymacji przedstawiono w tabeli 2.

Wartość prawdopodobieństwa testowego oraz współczynnika determinacji $R^2 = 80\%$ pozwala uznać zaproponowany poniższy model prognozowania popytu za ostateczny. Ma on następującą postać (1):

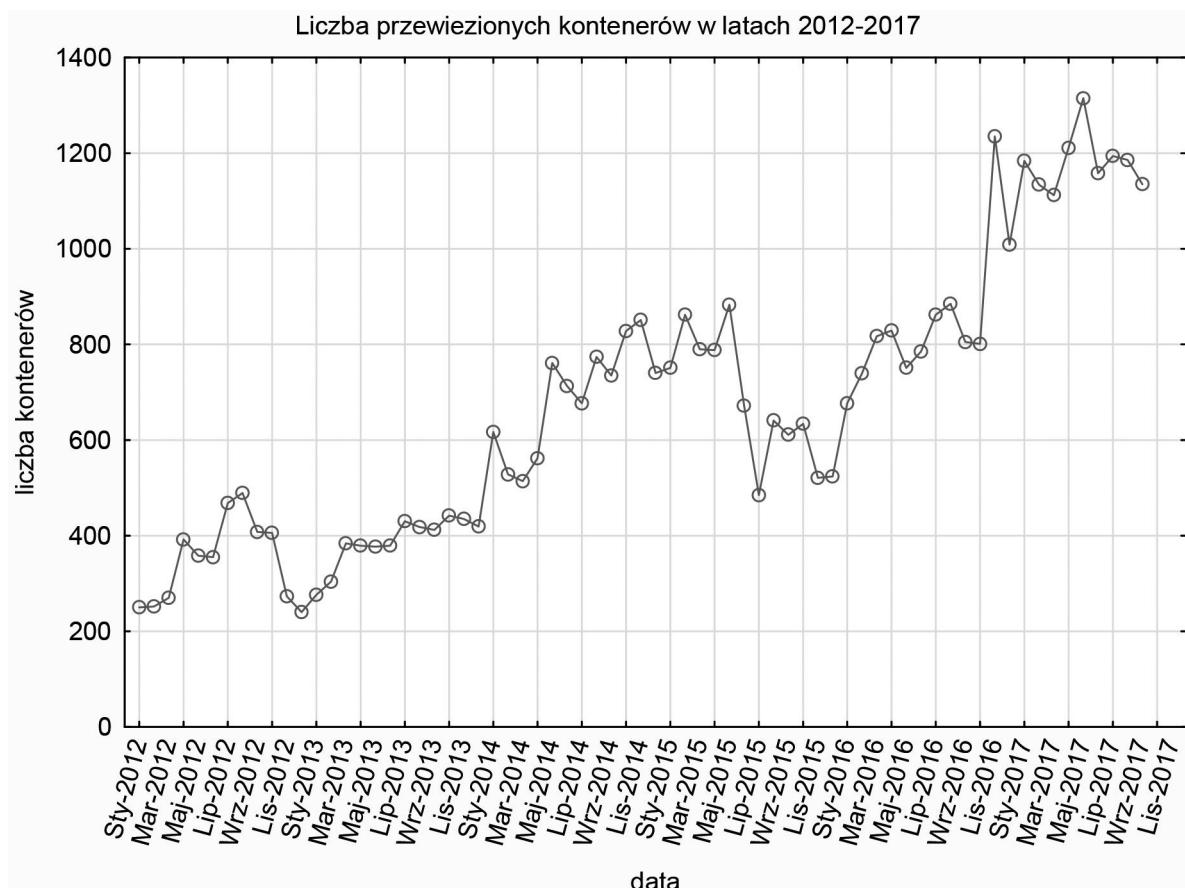
$$\hat{y} = 0,18 * t^2 + 378 \quad (1)$$

Dopasowanie funkcji trendu do danych rzeczywistych przedstawia rysunek 2.

Kolejnym etapem jest analiza występujących odchyleń od trendu (Dittmann, Szabela-Pasierbińska, Dittmann, Szpulak, 2011). Mogą być one sezonowe, w tym przypadku miesięczne lub/i losowe. Potwierdzenie istnienia wahań sezonowych następuje

Rysunek 1

Wykres przebiegu analizowanej zmiennej



Źródło: opracowanie własne na podstawie danych udostępnionych przez badaną firmę.

Tabela 1

Podsumowanie wyników regresji dla t i t^2 (parametry statystycznie istotne zostały podkreślone)

N = 69	R = 0,9029 R ² = 0,8153 Popraw. R2 = 0,8098 F(2,66) = 145,76					
	b*	Bł. std.z b*	b	Bł. std.z b	t(66)	p
W. wolny			301,4490	46,84734	6,434709	0,000000
t	0,401642	0,214511	5,7826	3,08840	1,872357	0,065588
t^2	0,508301	0,214511	0,1013	0,04276	2,369578	0,020741

Źródło: opracowanie własne.

Tabela 2

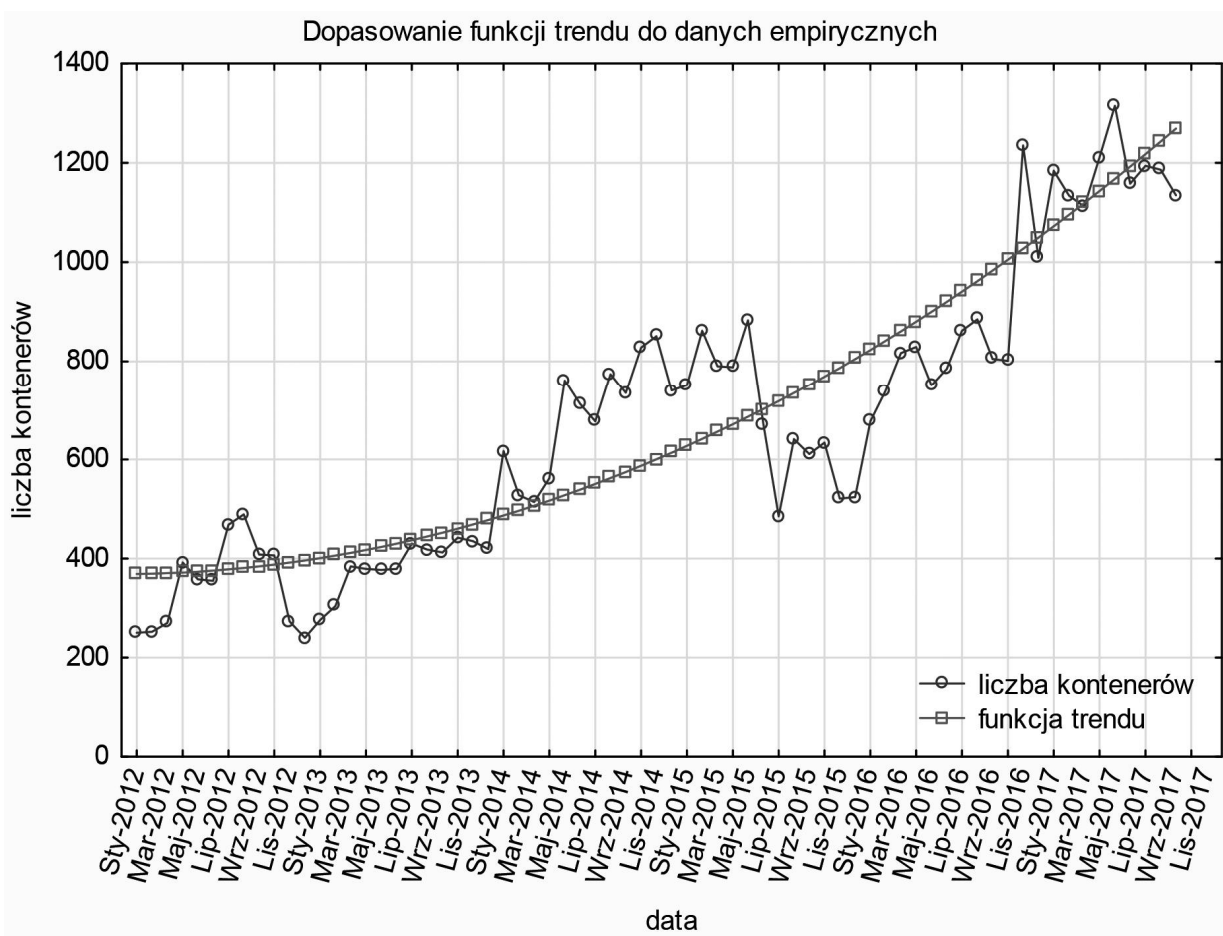
Podsumowanie wyników regresji dla t^2 (parametry statystycznie istotne zostały podkreślone)

N = 69	R = 0,8975 R ² = 0,8056 Popraw. R2 = 0,8027 F(1,67) = 277,62		
	b	Bł. std. b	p
W. wolny	378,0196	23,27615	0,000000
t^2	0,1789	0,01074	0,000000

Źródło: opracowanie własne.

Rysunek 2

Dopasowanie funkcji trendu do danych empirycznych



Źródło: opracowanie własne.

w drodze testowania istotności różnic średnich czasów dla poszczególnych miesięcy. Pomocna jest w tym analiza wariancji ANOVA, jednak jej przeprowadzenie wymaga spełnienia założenia normalności rozkładów we wszystkich analizowanych zbiorowościach, w tym przypadku miesiącach. Ponadto powinny one mieć jednakową wariancję (StatSoft. 2006). W związku z powyższym za pomocą testu Shapiro-Wilka zbadano zgodność rozkładu zmiennej w grupach dla wszystkich miesięcy. Otrzymane wyniki prawdopodobieństwa *p-value* przedstawia tabela 3.

Tabela 3
Wyniki testu Shapiro-Wilka

Lp.	Miesiąc	Poziom prawdopodobieństwa <i>p</i>
1	styczeń	0,00753
2	luty	0,34036
3	marzec	0,71534
4	kwiecień	0,58469
5	maj	0,20945
6	czerwiec	0,19279
7	lipiec	0,67102
8	sierpień	0,19614
9	wrzesień	0,42427
10	październik	0,55261
11	listopad	0,40692
12	grudzień	0,98874

Źródło: opracowanie własne.

Dla wszystkich miesięcy, z wyjątkiem stycznia, wartość statystyki testowej nie pozwala na odrzucenie hipotezy H_0 informującej o tym, że rozkład badanej zmiennej jest zbliżony do normalnego, o czym świadczą przedstawione powyżej wartości prawdopodobieństw testowych *p*. Następnie dokonano sprawdzenia jednorodności wariancji, wykorzystując w tym celu test Levene'a, dla którego uzyskano wartość $p = 0,9911$, oraz test Browna-Forsythe'a z prawdopodobieństwem $p = 0,9982$. Oba testy zakładają w hipotezie H_0 , potwierdzoną otrzymanym wynikiem, jednorodność wariancji pomiędzy badanymi próbkami. Pewna odporność testu ANOVA na odstępstwo od normalności oraz potwierdzona jednorodność wariancji pozwoliły na zastosowanie analizy celem

Tabela 4
Wyniki analizy jednorodności wariancji ANOVA

Zmienna	Test jednorodności wariancji							
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
eksport	67259,61	11	6114,510	1954382	57	34287,39	0,178331	0,998203

Źródło: opracowanie własne.

sprawdzenia równości średnich w grupach (Sokołowski, 2016; StatSoft. 2006). Jej wyniki przedstawiono w tabeli 4.

Zgodnie z przytoczoną analizą (wartość $p = 0,9982$) nie ma podstaw do odrzucenia hipotezy zerowej, zgodnie z którą wszystkie wartości przeciętne w grupach są takie same. W związku z powyższym w badanym zjawisku nie występują wahania sezonowe, lecz jedynie losowe i ostateczny model będzie miał następującą postać (2):

$$y_t = f(t) + \xi_t \quad (2)$$

gdzie:

$f(t)$ — wyznaczona już wyżej funkcja trendu,

ξ_t — wahania losowe.

Przyjęty model nie oddaje w pełni charakterystyki zjawiska, o czym świadczy uzyskany współczynnik determinacji $R^2 = 80\%$. Wprawdzie diagnostyka rozkładu jego reszt wskazuje, że mają one rozkład normalny co pokazano na rys. 3. Świadczy o tym wartość prawdopodobieństwa testowego testu Shapiro-Wilka wynosząca $p = 0,59$, jednak badanie funkcji autokorelacji i autokorelacji cząstkowej wskazuje, że istnieją zależności, które nie zostały wyjaśnione przyjętym modelem (Pawełek, Wanat, Zeliaś, 2013). Stało się to powodem podjęcia próby jego ulepszenia.

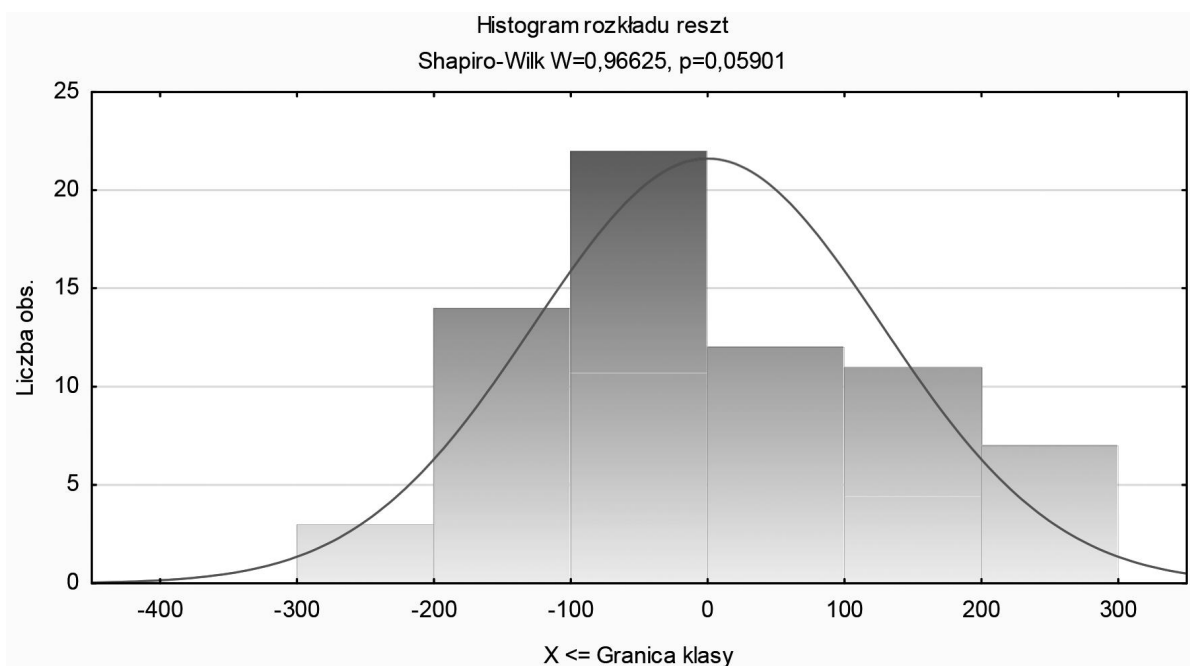
Model ARIMA

W celu zwiększenia jakości predykcji zaproponowano model ARIMA. Jest to model, w którym wartość szeregu jest skorelowana zarówno ze swoją poprzednią wartością [składowa autoregresyjna AR (*p*)], jak i ze składnikiem losowym [składowa średniej ruchomej MA (*q*)]. Łącznik I oznacza, że szereg został poddany operacji *d*-krotnego różnicowania. Stosowanie modelu ARIMA jest jednak ograniczone wymaganiami dotyczącymi stacjonarności badanego procesu (Bielińska, 2007). Wykres autokorelacji zmiennej wskazuje na niespełnienie tego warunku (rys. 4), dlatego konieczne było usunięcie niestacjonarności za pomocą różnicowania.

Otrzymany szereg, o stałych momentach pierwszego i drugiego rzędu, jest bazą do ustalenia paramet-

Rysunek 3

Histogram rozkładu reszt



Źródło: opracowanie własne.

trów modelu ARIMA (Bielińska, 2007; Dittmann, 2000; Sokołowski, 2016), do których należą wspomniane wyżej:

- parametr p , będący składową opisującą autoregresyjny element modelu ARIMA;
- parametr q , stanowiący składową związaną z elementem autoregresyjnym;
- parametr d , mówiący o rzędzie różnicowania, został ustalony w procesie osiągania stacjonarności i dla analizowanego procesu wynosi 1.

Narzędziem wspierającym proces estymacji jest przebieg wykresów funkcji autokorelacji ACF i autokorelacji cząstkowej PACF po zróżnicowaniu (rys. 5 oraz rys. 6).

Ich przebieg wskazuje, że analizowany proces jest autoregresyjny, rzędu maksymalnie drugiego (na podstawie PACF) oraz że występuje silna ujemna autokorelacja rzędu pierwszego, a pozostałe nie są istotne (ACF). W związku z powyższym w pierwszej kolejności zaproponowano model ARIMA (2,1,0). Ponieważ analiza autokorelogramów nie zawsze daje jednoznaczne wyniki i ważne są także inne czynniki weryfikujące model, jak kryteria informacyjne, liczba parametrów użytych do modelu (dążymy do budowy modelu o jak najmniejszej liczbie) oraz ich statystyczna istotność, estymowano także inne modele, przedstawione w tabeli 5.

Dla modeli ARIMA (0,1,1), ARIMA (1,1,0), ARIMA (2,1,0) wszystkie oszacowane parametry

były statystycznie istotne, jednak najniższe wartości kryteriów informacyjnych oraz średniego błędu estymacji SE (ang. *standard error*) uzyskano (zgodnie z wcześniejszymi przypuszczeniami) dla modelu ARIMA (2,1,0). Dlatego właśnie ten model poddano dalszej diagnostyce, sprawdzając normalność rozkładu reszt oraz istnienie statystycznie istotnych autokorelacji. Rozkład okazał się zbliżony do normalnego, a wykreślone autokorelogramy (rys. 7 oraz rys. 8) zaprezentowały brak istotnych wartości funkcji.

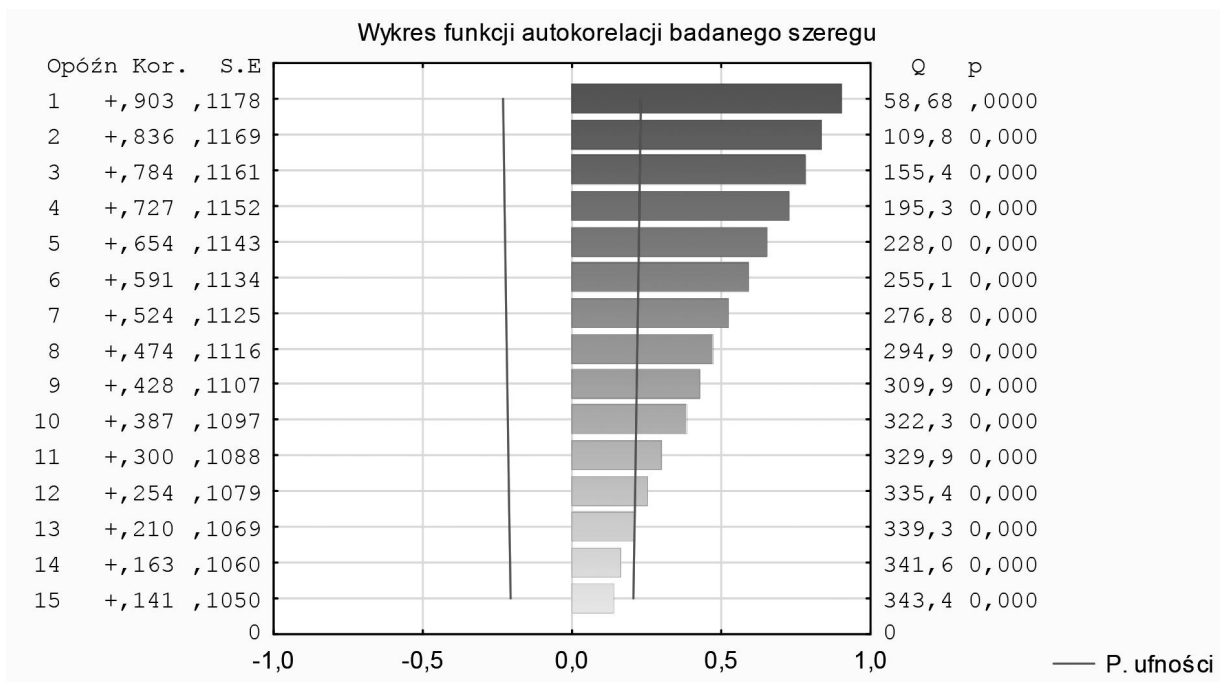
Poprawność modelu pozwoliła na wyznaczenie wartości prognozy popytu. Wykres wartości empirycznych i prognozy przedstawia rysunek 9.

O jakości modelu ARIMA (2,1,0) ostatecznie zdecydowało porównanie z empirycznymi obserwacjami testowymi, które nie wzięły udziału w budowie modelu. W tabeli 6 zestawiono wyniki dla dwóch zaproponowanych modeli i porównano z danymi testowymi.

Okazuje się, że dużo lepiej z opisaniem ewolucji popytu poradził sobie model regresji, uzyskując satysfakcjonujące błędy prognoz na poziomie nieprzekraczającym 3%. Predykcje według modelu ARIMA (2,1,0) okazały się zdecydowanie gorsze, jednak nie przekroczyły poziomu 10%, co w przypadku jakościowych analiz popytu jest również korzystne. Średnie błędy prognoz przedstawiono w tabeli 7.

Rysunek 4

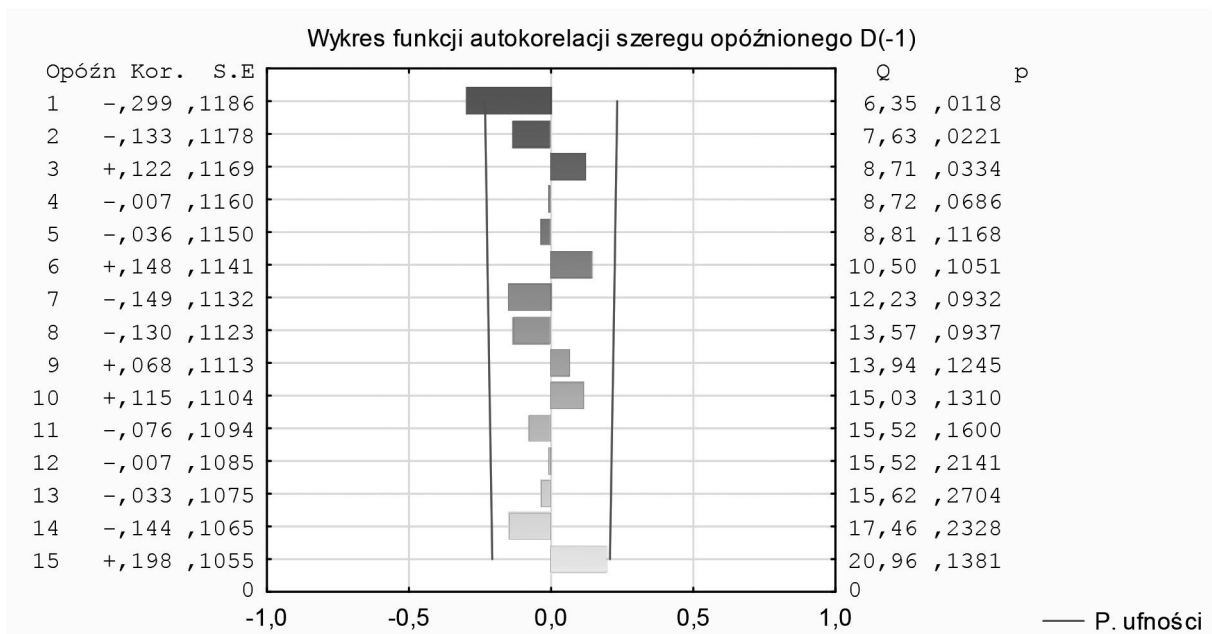
Wykres funkcji autokorelacji badanego szeregu



Źródło: opracowanie własne.

Rysunek 5

Wykres funkcji autokorelacji szeregu opóźnionego



Źródło: opracowanie własne.

Tabela 5

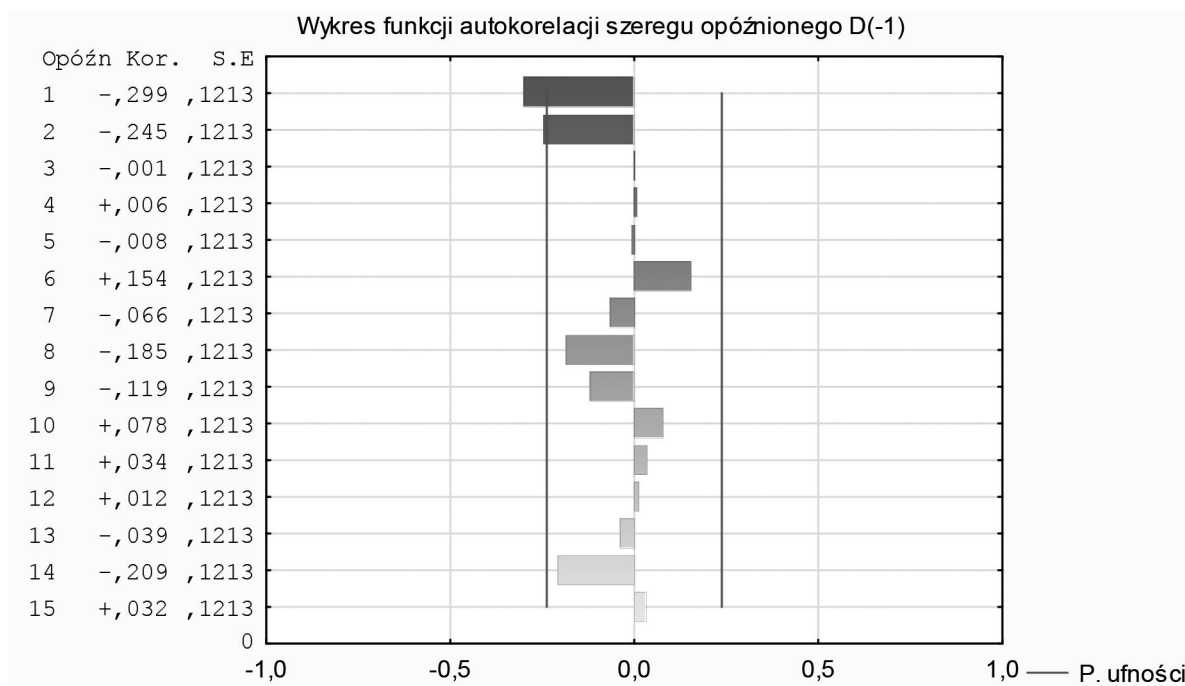
Wyniki estymacji zaproponowanych modeli ARIMA (parametry statystycznie istotne zostały podkreślone)

Model	ARIMA (0,1,1)	ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (1,1,1)
Przekształcenia:	D (1)	D (1)	D (1)	D (1)
Stała	13,455	13,263	13,467	13,475
SE	7,0417	9,2702	7,2783	6,8887
q(1)	0,40658			0,47455
SE	0,11465			0,27016
q(2)				
SE				
p(1)		-0,3006	-0,3757	0,08712
SE		0,11868	0,12161	0,23198
p(2)			-0,2481	
SE			0,12165	
MS	9264,2	9676,8	9226,3	9390,9
Kryt. inform. Akaike'a	818,2291	821,1067	818,1498	820,1104
Kryt. bayes. Schwarz	824,8877	827,7652	824,8279	828,9884
Kryt. Hannana-Quinna	820,8674	823,7450	820,4676	823,6281

Źródło: opracowanie własne.

Rysunek 6

Wykres funkcji autokorelacji cząstkowej szeregu opóźnionego



Źródło: opracowanie własne.

Podsumowanie

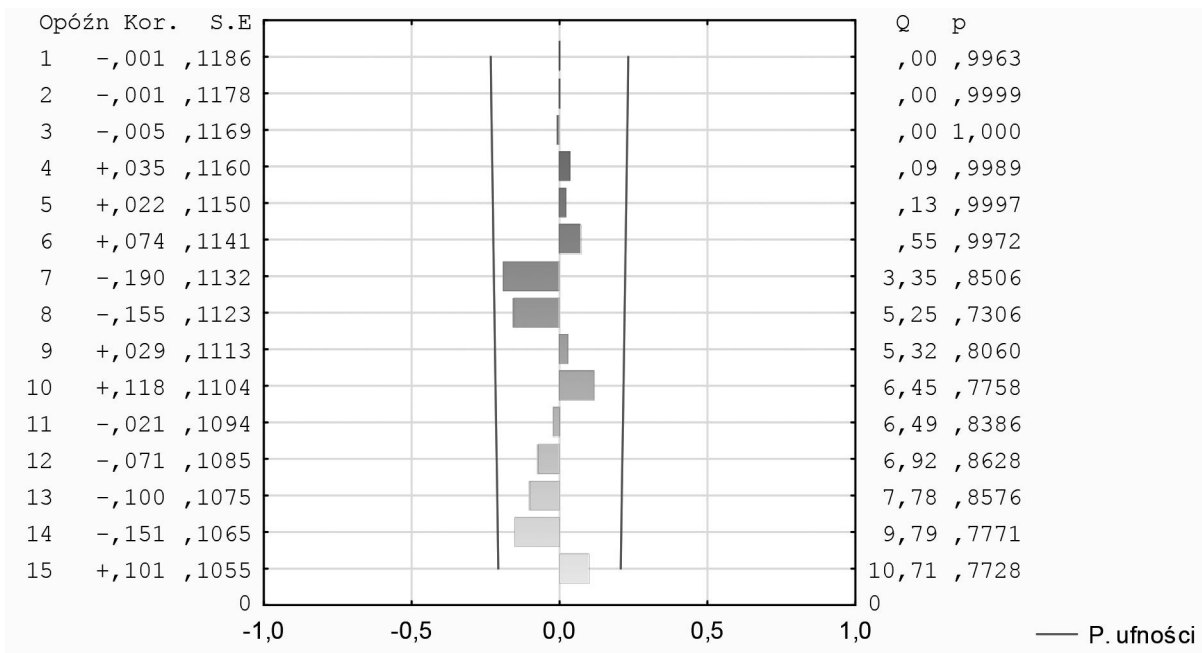
Założeniem badania było uzyskanie wiarygodnego modelu opisującego ewolucję popytu na eksport kontenerów, która pozwoliłaby określić potencjalne kierunki dalszego działania. Prognozowana liczba zamówień rośnie w szybkim tempie, co wpisuje się w ogólnoświatowe trendy sprzyjające rozwojowi

transportu intermodalnego. Dla analizowanej firmy jest to z jednej strony potwierdzenie, że przyjęta strategia była słuszna, a z drugiej wskazanie perspektywy ekspansji w tej dziedzinie, np. poprzez zwiększenie liczby połączeń lub rozbudowę już istniejących tras.

Opracowane modele dobrze poradziły sobie w predykcji, chociaż dużo lepsze wyniki uzyskano dla prostego modelu regresji. Nie stwierdzono jednak

Rysunek 7

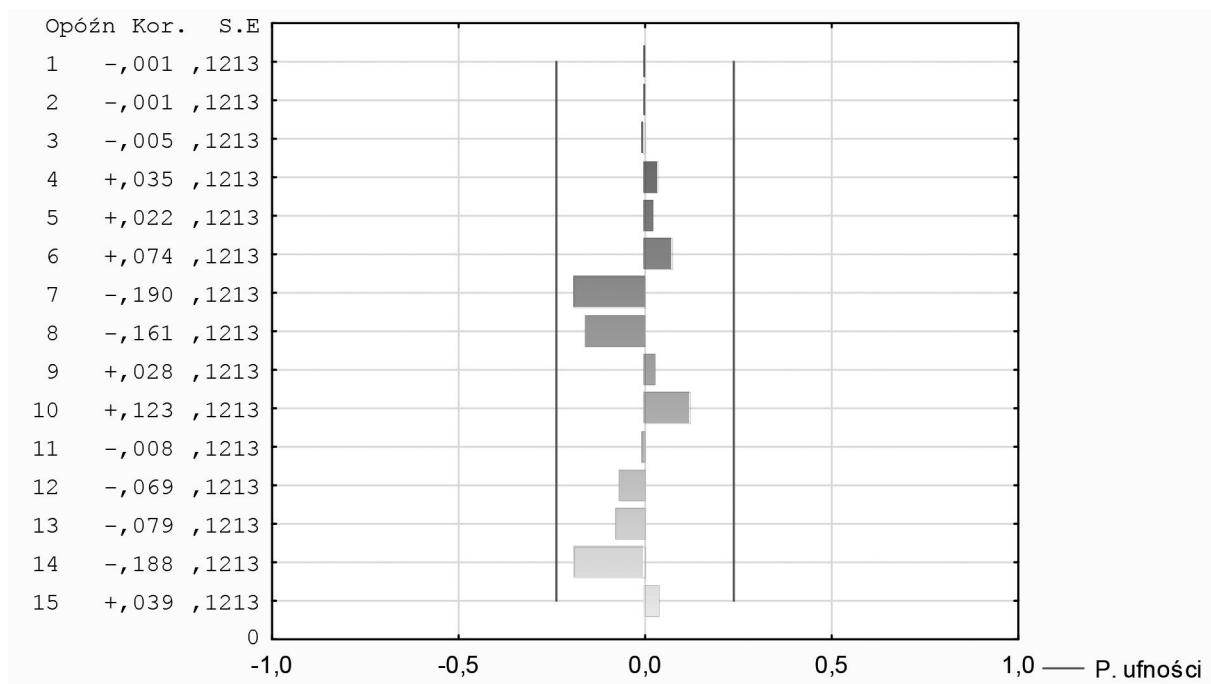
Funkcja autokorelacji reszt modelu ARIMA (2,1,0)



Źródło: opracowanie własne.

Rysunek 8

Funkcja autokorelacji cząstkowej reszt modelu ARIMA (2,1,0)



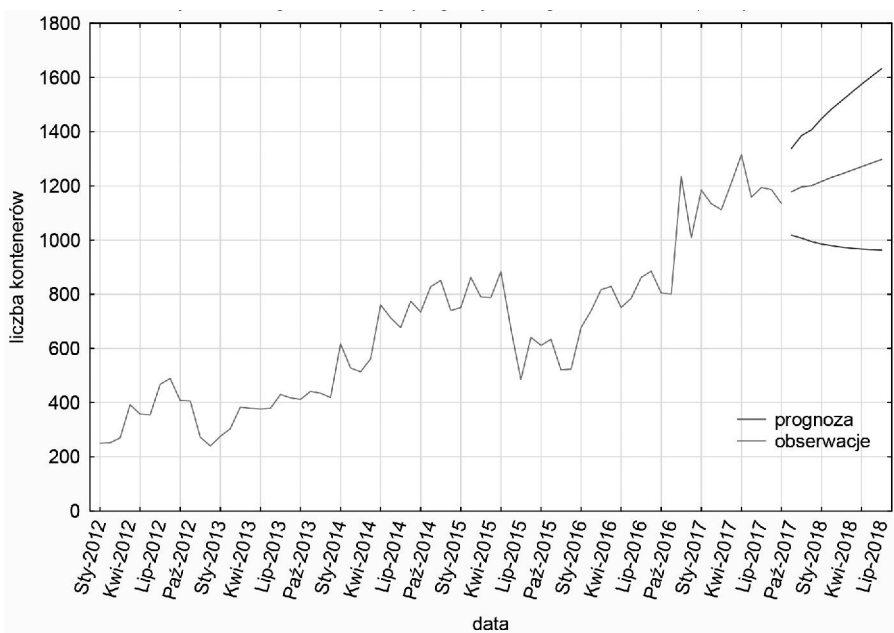
Źródło: opracowanie własne.

zakładanej pierwotnie sezonowości procesu, co może świadczyć o racjonalnym zarządzaniu przepływami i ich skutecznej optymalizacji. Dalszym kierunkiem

badania mogłaby być analiza mniej zagregowanych danych i wnioskowanie na temat tygodniowych bądź dziennych przebiegów pociągów.

Rysunek 9

Wykres wartości empirycznych i funkcji prognozy według modelu ARIMA (2,1,0)



Źródło: opracowanie własne.

Tabela 6

Porównanie wyników prognoz wg modelu regresji oraz modelu ARIMA (2,1,0)

Rok	Miesiąc	Prognoza model regresji	Prognoza model ARIMA (2,1,0)	Dane empiryczne	Względny błąd prognozy Ψ (%) regresja	Względny błąd prognozy Ψ (%) ARIMA (2,1,0)
2017	10	1254,30	1178,01	1215	-3,23	3,04
2017	11	1278,81	1196,37	1280	0,09	6,53
2017	12	1303,32	1200,67	1290	-1,03	6,92
2018	1	1327,83	1216,37	1320	-0,59	7,85
2018	2	1352,34	1231,27	1352	-0,03	8,93
2018	3	1376,85	1243,64	1360	-1,24	8,56

Źródło: opracowanie własne.

Tabela 7

Względne średnie błędy prognoz dla zaproponowanych modeli

	Względny błąd prognozy Ψ (%) regresja	Względny prognozy Ψ (%) błąd ARIMA (2,1,0)
Średni względny błąd prognozy	1,04	6,97

Źródło: opracowanie własne.

Bibliografia

- Bielińska, E. (2007). *Prognozowanie ciągów czasowych*. Gliwice: Wydawnictwo Politechniki Śląskiej.
- Dittmann, P. (2000). *Metody prognozowania sprzedaży w przedsiębiorstwie*. Wrocław: Wydawnictwo Akademii Ekonomicznej.
- Sokołowski, A. (2016). *Prognozowanie i analiza szeregów czasowych. Materiały szkoleniowe*. Kraków: StatSoft Polska.
- Czyżycki, R., Klóska, R. (2011). *Ekonometria i prognozowanie zjawisk ekonomicznych w przykładach i zadaniach*. Szczecin: Economicus.
- Dittmann, P., Szabela-Pasierbińska, E., Dittmann, I., Szpulak, A. (2011). *Prognozowanie w zarządzaniu sprzedażą i finansami przedsiębiorstwa*. Warszawa: Wolters Kluwer Polska Sp. z o.o.
- Pawełek, B., Wanat, S., Zeliaś, A. (2013). *Prognozowanie ekonomiczne Teoria przykłady zadania*. Warszawa: PWN.
- StatSoft. (2006). *Elektroniczny Podręcznik Statystyki*. Kraków.
- <http://www.statsoft.pl/textbook/stathome.html> (28.02.2018).