

Uniwersytet Warszawski

Wydział Nauk Ekonomicznych

Ewa Rudzińska

432501

Weronika Wyrwas

432578

Prognozowanie cen ziemniaków w oparciu o różne metody prognostyczne, przypadek Polski.

Praca zaliczeniowa z przedmiotu:

Prognozowanie i symulacje

wykonana pod kierunkiem

dr. Łukasza Postka

Warszawa, czerwiec 2025

Prognozowanie i symulacje
Projekt ekonometryczny

<i>Autor/Autorzy</i>	Ewa Rudzińska, 432501 Weronika Wyrwas,
<i>Tytuł</i>	Prognozowanie ceny ziemniaków w oparciu o różne metody prognostyczne
<i>Typ projektu</i>	prognozowanie
<i>Wykorzystanie SI</i>	<i>Podczas przygotowania tej pracy autor/autorzy nie użyli żadnego narzędzia SI.</i>

Informacje o artykule będącym inspiracją dla badania

<i>Tytuł</i>	Potato price forecasting with Holt-Winters and ARIMA methods: a case study
<i>Autor/Autorzy</i>	Mehmet Arif Sahinli
<i>Journal/Miejsce publikacji</i>	American Journal of Potato Research
<i>Rok</i>	2020
<i>Zakres stron</i>	1-11
<i>Tematyka, problemy i cele badawcze</i>	Artykuł poświęcony był prognozowaniu cen konsumpcyjnych ziemniaków w Turcji. Autorzy zanaznaczają że cena ziemniaka ustalana jest w wyniku wolnego rynku, co wiąże się z występowaniem niepewności w kontekście producentów. Dodatkowo zwracają oni uwagę na brak skutecznej organizacji odpowiedzialnej za produkcję i marketing ziemniaków. Wynikiem tego są wahania cen oddziałujące nie tylko na konsumentów, ale również na cały przemysł spożywczy. Głównym celem analizy było porównanie skuteczności w obliczu analizowanego problemu dla trzech metod: modelu Holta-Wintersa w wersji addytywnej i mnożystycznej oraz modelu ARIMA.
<i>Główne wnioski</i>	W badaniu przetestowano różne wartości początkowe dla metody Holta-Wintersa, jak się okazało model w postaci addytywnej dawał bardziej dokładne wyniki prognozowania. W przypadku modelu ARIMA, został on uznany za niestacjonarny, dlatego policzone zostały pierwsze różnice, co pozwoliło na spełnienie założenia o stacjonarności. W kolejnym kroku na podstawie korelogramów ACF i PACF wyznaczone zostały komponenty p i q. W wyniku tej analizy wyznaczony został model ARIMA (1,1,2), jednak w dalszej części

	pracy przetestowane zostały również inne kombinacje tego modelu. Ostatecznie model ARIMA (1,1,2) okazał się być najlepszym możliwym modelem w odniesieniu do kryteriów informacyjnych AIC i BIC, dodatkowo spełniając założenie o tym, że reszty mają być białym szumem. Ostateczna analiza błędów prognoz w uparciu o RMSE, MAE i MAD wykazała, że model ARIMA (1,1,2) okazał się być najlepszym analizowanym modelem, wnosząc tym samy pewien wkład dla ustalania strategii przez tureckich ekonomistów.
<i>Metodyka badawcza</i>	W artykule bazowym wykorzystano metody Holta-Wintersa w wersji addytywnej oraz multiplikatywnej, dodatkowo rozszerzając analizę o metodę Boxa-Jenkinsa ARIMA. Do przeprowadzenia tego badania wykorzystano pakiet Eviews 10. Celem było zbadanie wpływu wartości początkowych trenu na dokładność prognozowania oraz znalezienie wyników które będą najlepiej dopasowane. Natomiast dokładność prognoz oceniono w oparciu o RMSE, MAE i MAD. Dodatkowo próba została podzielona na dwie części: estymacyjną (styczeń 2005 - lipiec 2019) oraz progностyczną (sierpień 2019 – grudzień 2019).
<i>Dane</i>	Dane pochodzą z Turkish Statistical Institute (Turkstat), obejmując indeksy cen konsumpcyjnych (CPI) oraz ceny towarów. W analizie wykorzystano dane miesięczne cen konsumpcyjnych ziemniaków w Turcji od 1 stycznia 2005 do 31 lipca 2019 roku.
<i>Dlaczego wybrano właśnie ten artykuł?</i>	Artykuł stanowi doskonały punkt odniesienia do przeprowadzenia podobnej analizy dla Polski. W jasny sposób opisuje całą metodologię zawierając przy tym szczegółowy opis techniczny metod i ich implementacji.

Podstawowe informacje o badaniu

<i>Problem badawczy</i>	Badnie poświęcone jest prognozowaniu cen ziemniaków w Polsce. Głównym celem jest zastosowanie metod prognostycznych wykorzystanych w artykule bazowym, rozrzeszając go jednak o inne metody oraz testy diagnostyczne, podejmując decyzje o ostatecznym wyborze najlepszego modelu w oparciu o analizę błędów prognoz. Analiza pozwoli na weryfikację czy proste metody radzą sobie równie dobrze jak bardziej zaawansowane, pozwalając tym samym na odniesienie otrzymanych wyników do tych uzyskanych dla analizy przeprowadzonej na innym rynku.
-------------------------	--

<i>Metodyka badawcza</i>	W badaniu wykorzystane zostały metody takie jak: metoda naiwna, średnia ruchoma, proste wygładzanie wykładnicze, model Holt-Wintersa w wersji addytywnej i multiplikatywnej, ETS (automatyczny dobór modelu) oraz model SARIMA. Ze względu na dostępność danych, zawierających sezonowość wykorzystana została metoda Tramo-Seats. Dodatkowo przeanalizowane zostały wyniki testu Diebolda-Mariano dla par prognoz. Ocena trafności prognoz oraz ostateczny wybór najlepszego modelu przeprowadzone zostały w oparciu o RMSE, MAE i MAPE. Cała analiza przeprowadzona została w środowisku R.
<i>Dane</i>	Dane wykorzystane w analizie pobrane zostały z GUS-u (Banku Danych Lokalnych). Dotyczą one przeciętnych cen detalicznych towarów i usług konsumpcyjnych, w analizowanym przypadku cen ziemniaków w zł/kg. Obejmują one okres od stycznia 2009 roku do grudnia 2018 roku. Szereg podzielony został na próbę estymacyjną oraz testową. Dodatkowo przeprowadzona została dekompozycja szeregu, w celu identyfikacji struktury szeregu.
<i>Główne wnioski</i>	Przeprowadzona analiza wykazała na występującą przewagę prostych metod prognostycznych, nad bardziej zaawansowanymi metodami. Najlepszą z metod okazała się średnia ruchoma, dla optymalnego okna równego 3 miesiące z RMSE równym 0.0737, natomiast najgorszą z nich był model SARIMA. Może to zatem skazywać, że dla rynku ziemniaków charakteryzującego się wysoką zmiennością, jak również szybką reakcję na zmiany rynkowe, to proste metody bazujące na ostatnich obserwacjach są dużo bardziej skuteczne, niż metody wykorzystujące długookresowe wzorce.

Ekonometryczny opis badania

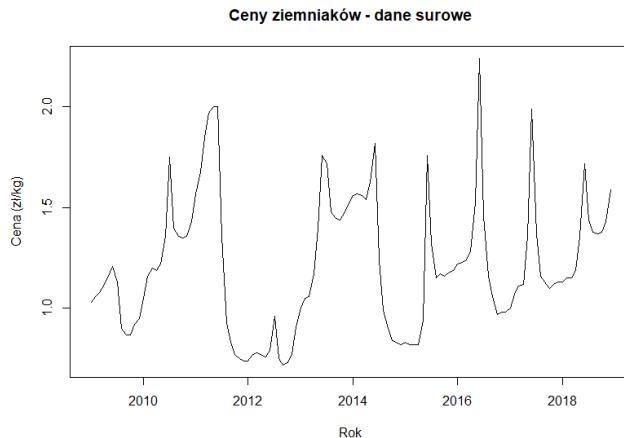
Historia ziemniaka w Polsce sięga 1683 roku, kiedy to Jan III Sobieski przywiózł z wyprawy wiedeńskiej jego bulwy jako podarunek dla królowej, by mogły one cieszyć oko w ogrodach wilanowskich. Z biegiem czasu zyskał on jednak na znaczeniu również wśród arystokracji jak i ludzi mniej zamożnych dominując wręcz polską kuchnię (Wojciech Nowacki, 2024). Polska jest jednym z największych producentów ziemniaka w Europie, plasując się na czwartym miejscu zaraz po Niemczech, Francji i Holandii. Jego zbiory w 2024 roku wynosiły 5.9 mln ton, stanowiąc bazę dla przemysłu spożywczego przy produkcji frytek, chipsów, skrobii ziemnaiczanej czy alkoholu (Ministerstwo Rolnictwa i Rozwoju Wsi, 2024). Rynek ten charakteryzuje się zatem występowaniem zmienności cen wynikających bezpośrednio z warunków pogodowych wpływających na urodzaj zbiorów, jak również ze względu na występowanie sezonowego koponenu tego surowca (ceny maleją w okolicach jesieni i rosną wraz z kolejnymi miesiącami roku magazynowego). W występujących warunkach niepewności producenci muszą jednak podejmować decyzje dotyczące planowanej wielkości uprawy, co jest nie tylko ryzykowne dla nich, ale także dla przetwórców oraz samych konsumentów. W tym kontekście podejmowanie problemu analizy ceny ziemniaków nabiera dużo większego znaczenia. Rzetelne prognozy mogą pomóc w podejmowaniu racjonalnych decyzji dla każdego z uczestników łańcucha dostaw. Dodatkowo przyczyniając się do kształtowania skutecznej polityki rolnej oraz systemów ostrzegania przed kryzysami żywonościowymi.

Niniejsze badanie skupia się zatem na wypracowaniu sensowych analitycznych narzędzi progностycznych dla polskiego rynku ziemniaka inspirowany pracą Mehmet Arif Sahinli, opisującego podobne badanie jednak dla rynku tureckiego. W poniższej pracy wykorzystane zostały metody takie jak: metoda naiwna, średnia ruchoma, proste wygładzanie wykładnicze, model Holt-Wintersa w wersji addytywnej i multiplikatywnej, ETS (automatyczny dobór modelu) oraz model SARIMA. Natomiast wybór najlepszej metody prognostycznej został oparty o wyniki RMSE, MAE i MAPE. Adaptacja i rozszerzenie metodologii z artykułu będącego inspiracją tej pracy na warunki polskiego rynku może przyczynić się do lepszego zrozumienia mechanizmów kształtowania się cen dostarczając praktycznych narzędzi wspomagających decyzje ekonomiczne w sektorze ziemniaczanym.

Wykorzystane dane

W pracy omówiony zostanie sezonowy szereg cen detalicznych ziemniaków w Polsce wyrażony z złotówkach na kilogram (zł/kg). Dane pobrane zostały ze strony Głównego Urzędu Statystycznego – Bank Danych Lokalnych i dotyczą okresu od stycznia 2009 roku do grudnia 2018 roku, zawierając tym samym 120 obserwacji o charakterze miesięcznym. Badany szereg przedstawiony został na wykresie 1. Analiza wizualna wskazuje zatem na występujące cykliczne wahania sezonowe o charakterze rocznym. Widoczny jest wzorzec, w którym ceny rosną w miesiącach zimowo-wiosennych, co wiąże się z okresem magazynowym. Następnie ceny maleją osiągając swoje minimum w okolicy jesieni, co jest związane z okresem zbiorów i napływem świeżych ziemniaków na rynek.

Wykres 1. Ceny detaliczne ziemniaków w Polsce w latach 2009-2018

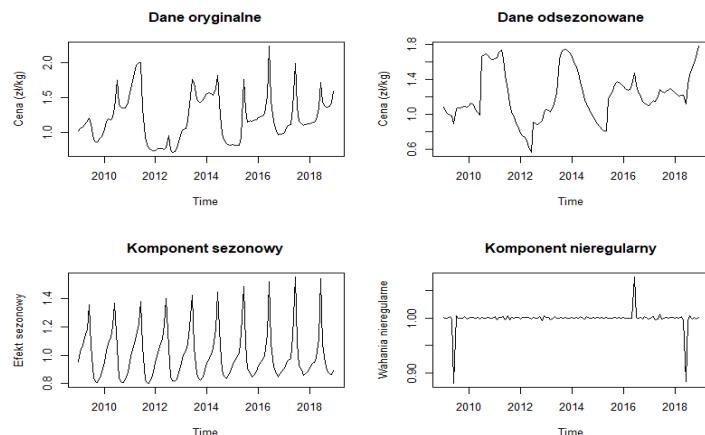


Źródło: Opracowanie własne w programie RStudio

Dekompozycja szeregu

Dekompozycja szeregu czasowego przeprowadzona została za pomocą metody Tramo-Seats w celu dogłębnej analizy struktury badanego szeregu, jak i określenia charakteru sezonowości oraz wykrycia obserwacji nietypowych. Umożliwiło to zatem na uzyskanie czterech komponentów: składnika trendu, składnika sezonowego, składnika nieregularnego oraz efektów kalendarzowych. Procedura ta automatycznie dobiera optymalny model SARIMA dla szeregu surowego, korygując o obserwacje odstające. Procedura ta pozwoliła zatem na uzyskanie zmiennej z usuniętą składową sezonową. Wyniki tego procesu widoczne są na wykresie 2. Analiza wykresu wskazuje, że sezonowość w badanym szeregu ma charakter multiplikatywny, co oznacza, że amplituda sezonowości zależy od poziomu szeregu. Mimo przeprowadzonej dekompozycji dalsza analiza opierać się będzie o dane surowe, przez wzgląd na przeprowadzanie rzetelnej analizy porównawczej wykorzystanych w dalszej części pracy metod, a sama dekompozycja służy jedynie celom diagnostycznym. Kolejnym krokiem podjętym w analizie było podzielenie próby na okres in-sample i out-of-sample. Okres out-of-sample obejmował dwanaście ostatnich miesięcy, czyli od stycznia do grudnia 2018 roku. Pozostałe 108 obserwacji uwzględnione zostały w próbce in-sample.

Wykres 2. Dekompozycja szeregu czasowego



Źródło: Opracowanie własne w programie RStudio

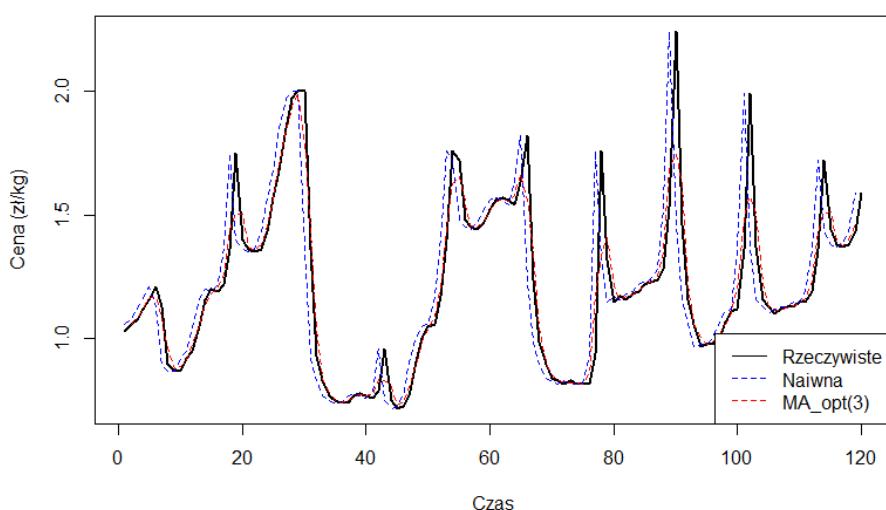
Prognoza naiwna

Prognoza naiwna stanowi jedną z najprostszych metod progностycznych, w których zakładamy że najbardziej prawdopodobną przyszłą wartością jest ostatnia obserwowana wartość. Pomiędzy swojej prostoty stanowi ona bardzo dobry benchmark dla bardziej zaawansowanych metod progностycznych. W przeprowadzonym badaniu wykorzystana została zmienna surowa, stanowiąc tym samym punkt odniesienia dla oceny względnej skuteczności pozostałych metod, której wyniki przedstawione zostały w tabeli porównawczej w końcowej części pracy.

Średnie ruchome

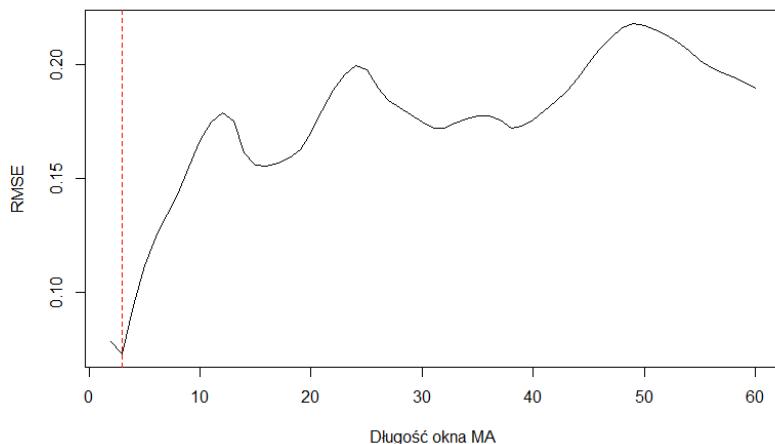
Prognoza naiwna wykorzystuje tylko ostatnią obserwację, przez co duży wpływ na prognozę ma komponent czysto losowy. Odpowiedzią na ten problem jest metoda średnich ruchomych, która zamiast korzystania z ostatniej obserwacji bazuje na k ostatnich obserwacjach wykorzystując średnią arytmetyczną. Ważne jest jednak ustalenie odpowiedniego okna prognozy. W pracy przedstawiono średnie ruchome różnych rzędów: 2, 3, 6 i 12 miesięcy oraz zastosowana została procedura optymalizacji okna, wybierając tą która minimalizuje błąd prognozy RMSE. Analiza wykazała że optymalne okno prognozy cen ziemniaków w Polsce wynosi 3 miesiące. Wizualizacja otrzymanych wyników przedstawiona została na wykresie 4. Widać że wraz ze wzrostem długości okna otrzymywane są coraz gorsze rezultaty dla prognozy krótkoterminowej, a optymalne okno prezentowane jest przez czerwoną linię. Wybór średniej z większego okresu przyczyniłby się zatem do otrzymania większego błędu prognozy. Przedstawiony został również wykres porównawczy dla danych rzeczywistych z prognozami uzyskanymi metodą naiwną oraz średnią ruchomą o optymalnych rzędzie MA(3). Widzimy że prognozy nie różnią się znacząco od siebie, natomiast prognoza uzyskana średnią ruchomą jest odróżnia gładszą i bardziej odporną na pojedyncze szoki. Obie metody radzą sobie jednak dobrze, co może sugerować że proste modele krótkoterminowe są wystarczające dla badanych danych.

Wykres 3. Analiza porównawcza rzeczywistych realizacji z metodą naiwną oraz średnią ruchomą dla optymalnego okna prognozy.



Źródło: Opracowanie własne w programie RStudio

Wykres 4. Zależność RMSE od optymalnego okna prognozy

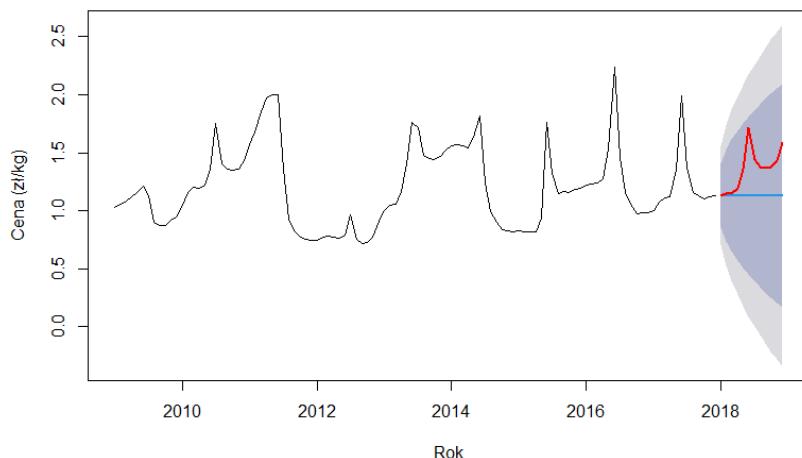


Źródło: Opracowanie własne w programie RStudio

Proste wygładzanie wykładnicze (SES)

Średnia ruchoma przypisuje identyczną wagę dla każdej z obserwacji. Odpowiedzią na to jest proste wygładzanie wykładnicze, które nadaje wyższą wagę dla nowszych obserwacji, przez co starsze obserwacje mają mniejszy wpływ na prognozę. W analizie wykorzystane zostały dane surowe. W pierwszym kroku przeprowadzona została estymacja na całym zbiorze danych w celu znalezienia parametrów wygładzania. W kolejnym kroku dla wyznaczenia rzetelnej prognozy analiza bazowała na ruchomym oknie. Wstępna analiza wyników wykazała, że parametr alpha wynosi 0.99, co jest bardzo bliskie 1, a zatem proste wygładzanie wykładnicze sprowadza się tak naprawdę do metody naiwnej. W kolejnym roku wyznaczone zostały błędy prognoz, które wykorzystane zostały w wyborze najlepszego modelu w sekcji końcowej. Dodatkowo przygotowana została wizualizacja jak dobrze SES radzi sobie z rzeczywistymi wahaniem cen, co widać na wykresie 5. Wyniki są zgodne z otrzymanym parametrem alpha. Model bardzo szybko zapomina, nie uwzględniając wzorców z przeszłości, działa wręcz jak prognoza naiwna.

Wykres 5. Prognoza prostego wygładzania wykładniczego w porównaniu do faktycznej realizacji.

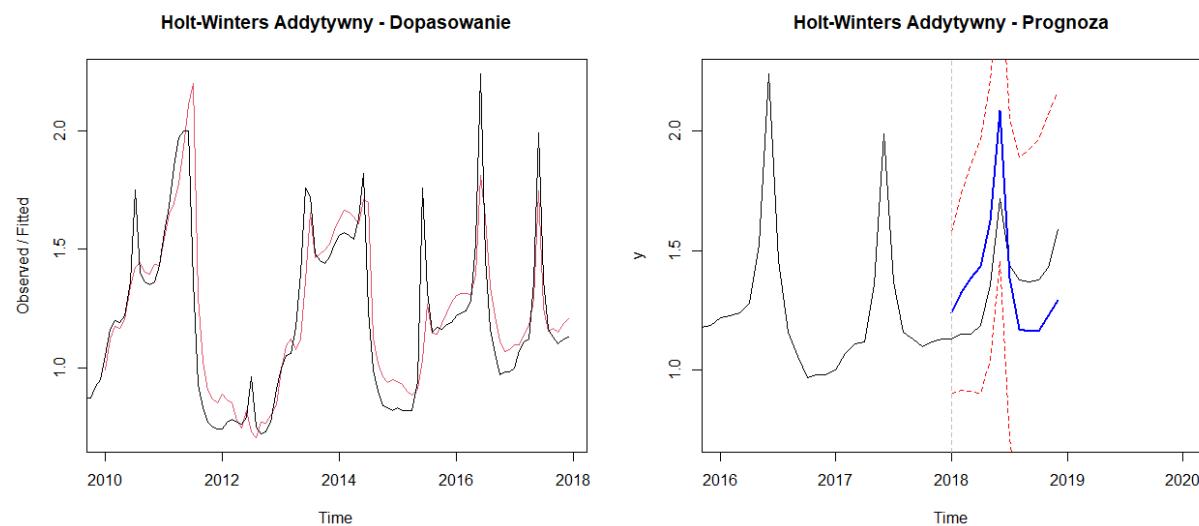


Źródło: Opracowanie własne w programie RStudio

Model Holta-Wintersa

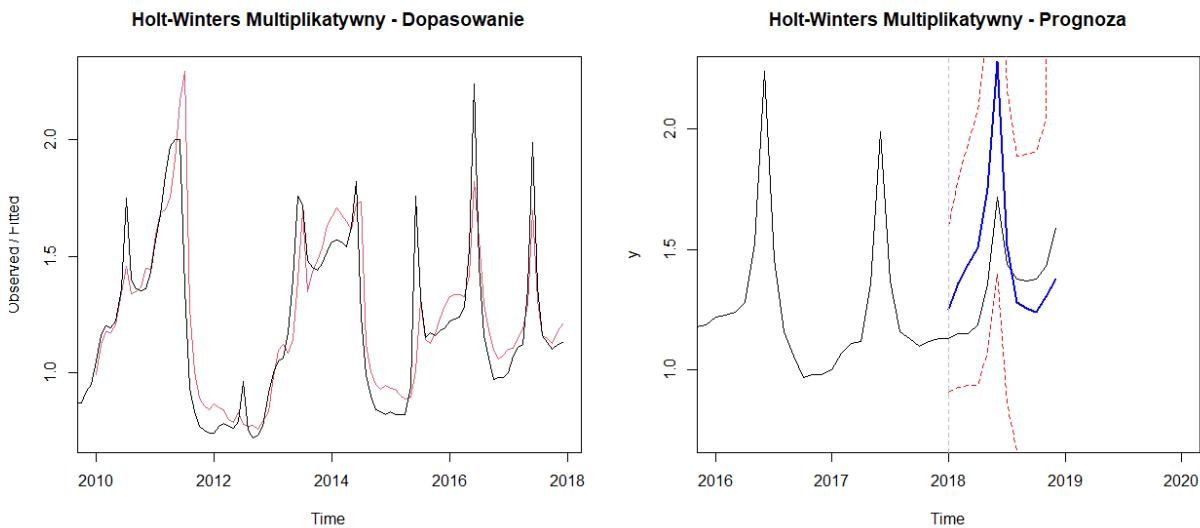
W kolejnym kroku przez wzgląd na występujące wzorce sezonowe do analizy wykorzystana została metoda Holta-Wintersa, która oprócz równania trendu uwzględnia również równanie sezonowości. Przez wzgląd na zastosowanie wersji addytywnej oraz multiplikatywnej w artykule bazowym obie wersje w tym przypadku również zostaną poddane analizie. Badanie przeprowadzone zostało danych surowych. Pierwszym krokiem było określenie próby in-sample i out-of-sample, które przyjmują identyczne przedziały opisane wcześniej. W kolejnym etapie, przy wykorzystaniu narzędzia RStudio wygenerowany został wykres modelu Holta-Wintersa dla modelu addytywnego cen ziemniaków w Polsce w latach 2009-2018, przedstawiony na wykresie 6. Lewa część przedstawia dopasowanie modelu Holta-Wintersa do analizowanych danych, natomiast po prawej stronie przedstawiona została prognoza oznaczona kolorem niebieskim z czerwonymi przedziałami ufności. W wyniku przeprowadzonej estymacji dla postaci addytywnej otrzymano wartości alpha równą 0.69, beta równą 0.008 oraz gamma wynoszącą 1. Identyczna wizualizacja wygenerowana została również dla modelu Holta-Wintersa w postaci multiplikatywnej, widoczna na wykresie 7. Wynik estymacji dla postaci multiplikatywnej wynoszą, dla alpha równą 0.7, beta równą 0 oraz gamma wynoszącą 1.

Wykres 6. Model Holta-Wintersa dla postaci addytywnej cen ziemniaków w Polsce w latach 2009-2018



Źródło: Opracowanie własne w programie RStudio

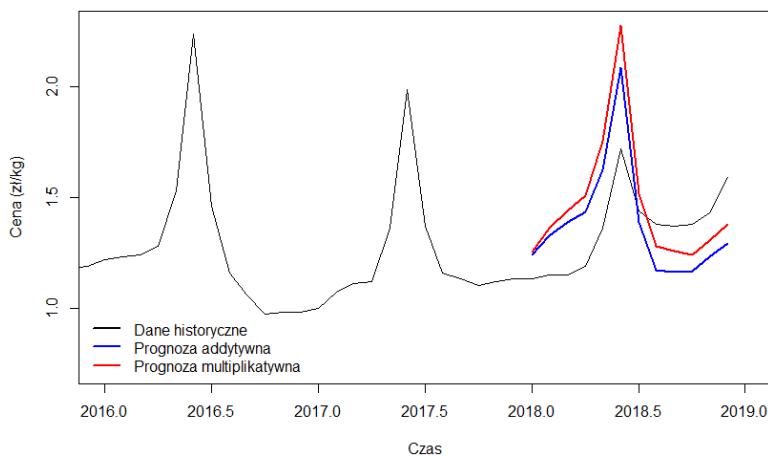
Wykres 7. Model Holta-Wintersa dla postaci addytywnej cen ziemniaków w Polsce w latach 2009-2018



Źródło: Opracowanie własne w programie RStudio

Porównując wyniki prognozy dla okresu out-of-sample dla postaci addytywnej oznaczonej kolorem niebieskim i postaci multiplikatywnej, oznaczonej kolorem niebieskim, widoczną na wykresie 8. z badanym szeregiem, można zauważyc że obie wersje dobrze odwzorowują sezonowość szeregu, nie można jednak jednoznacznie określić która z wersji jest lepsza. Dla okresu wzrostu cen postać addytywna trafniej uchwyciła skalę wzrostu, natomiast po osiągnięciu szczytu tu postać multiplikatywna lepiej odwzorowuje to, co jest możliwe do zaobserwowania przez dane historyczne. W tym przypadku przydatnym będzie przeanalizowanie wyników RMSE (szersza analiza pozostałych błędów prognoz zawarta w końcowej części pracy). RMSE dla postaci addytywnej wynosi 0.229, natomiast dla postaci multiplikatywnej wynosi 0.262, co sugeruje że postać addytywna modelu Holta-Wintersa jest dla powyższej analizy bardziej odpowiednia.

Wykres 8. Model Holta-Wintersa dla postaci multiplikatywnej oraz addytywnej cen ziemniaków w Polsce w latach 2009-2018

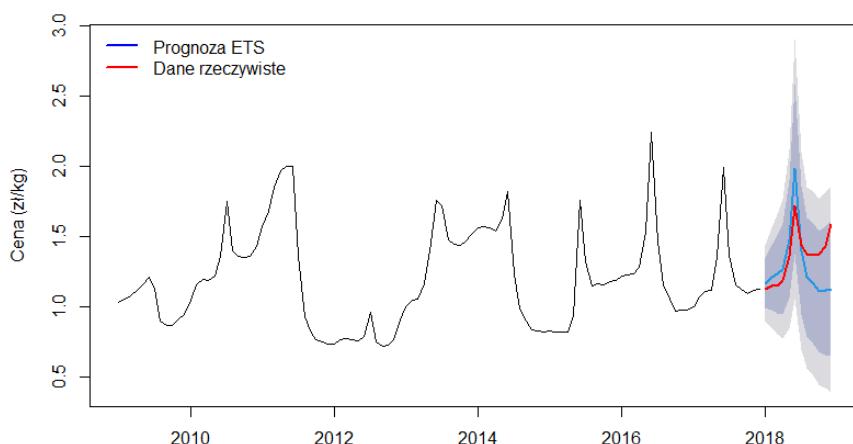


Źródło: Opracowanie własne w programie RStudio

Modele ETS

Modele ETS, są uproszczeniem modelu Holta-Wintersa, które pozwalają na uwzględnienie różnych kombinacji specyfiki wyróżnionych komponentów szeregu czasowego. Dla badanego zjawiska wykorzystana została automatyczna procedura, która wykazała, że komponent błędu jest mnożnikowy, tak samo jak komponent sezonowości, natomiast komponent trendu nie występuje. Parametry wygładzania dla tego modelu wynosiły: alpha równa 0.7697, co sugeruje że silnie reaguje on na ostatnie obserwacje, gamma równa 0.2303 wskazując na raczej umiarkowaną reakcję modelu na zmianę sezonowości. Wygenerowana została również prognoza w oparciu o model ETS, przedstawiona na wykresie 9. Analiza ostatnich 12 miesięcy pokazuje, że całkiem dobrze radzi sobie on z odwzorowaniem sezonowego wzorca, mieszcząc się w przedziale ufności, co może wskazywać na potencjalną konkurencyjność dla innych analizowanych metod.

Wykres 9. Model ETS dla cen ziemniaków w Polsce w latach 2009-2018



Źródło: Opracowanie własne w programie RStudio

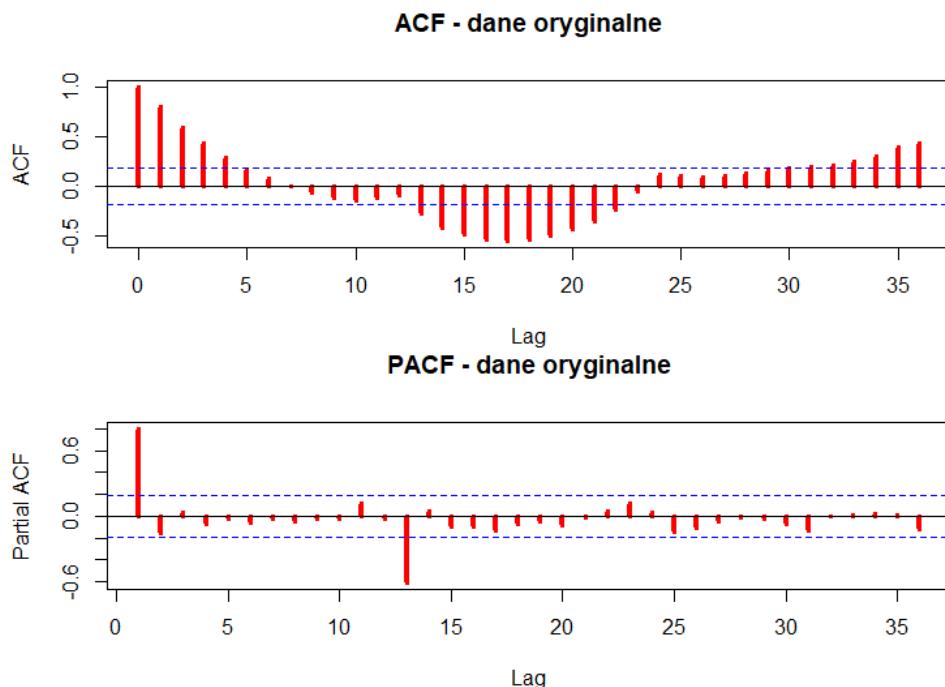
Model SARIMA

W kolejnej części pracy wykorzystany zostanie model SARIMA (Seasonal Autoregressive Integrated Moving Average), będący rozwinięciem standardowego modelu ARIMA (Autoregressive Integrated Moving Average), który uwzględnia sezonowość w analizie szeregów czasowych. Jest on szczególnie przydatny w przypadkach, gdy badany szereg wykazuje regularne sezonowe wzorce, będąc zatem jak najbardziej adekwatny dla analizy cen ziemniaków w Polsce, przez wzgląd na występujące wzorce sezonowe. Podobnie jak wcześniej próba została podzielona na okres in-sample oraz out-of-sample.

Warunkiem koniecznym pozwalającym na wykorzystanie tego modelu jest spełnienie kryterium stacjonarności szeregu. Wykres 1. przedstawiony na początku pracy sugeruje, że szereg nie jest stacjonarny, czego potwierdzenie można również znaleźć na wykresie 10. przedstawiającym ACF i PACF dla danych oryginalnych. ACF maleje powoli i ma wyraźny długi ogon, natomiast dla PACF pierwsze wartości są istotne statystycznie. Dlatego w kolejnym kroku nałożone zostały pierwsze róż-

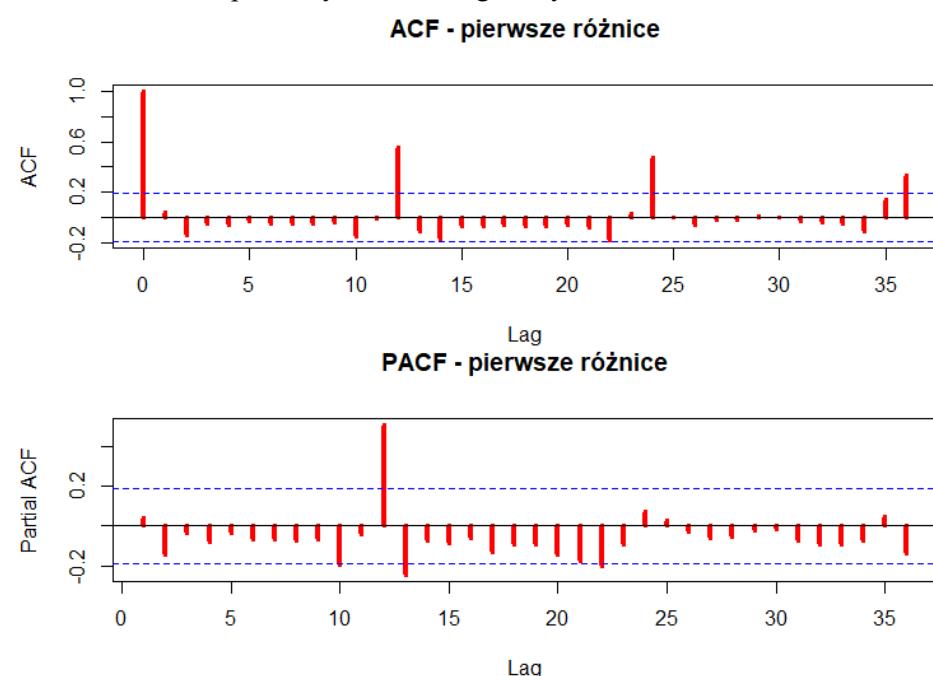
nice regularne. Tym razem analiza wykresu 11. ACF i PACF dla pierwszych różnic regularnych pokazuje, że obecne są jednak nadal korelacje sezonowe, co widać w przypadku ACF, ze względu na istotność statystyczną wypustek 12, 24, 36. Konieczne jest zatem nałożenie pierwszych różnic sezonowych, czego wynik przedstawiony został na wykresie 12. Zabieg ten powinien pozwolić na osiągnięcie stacjonarności.

Wykres 10. ACF i PACF dla cen ziemniaków w Polsce



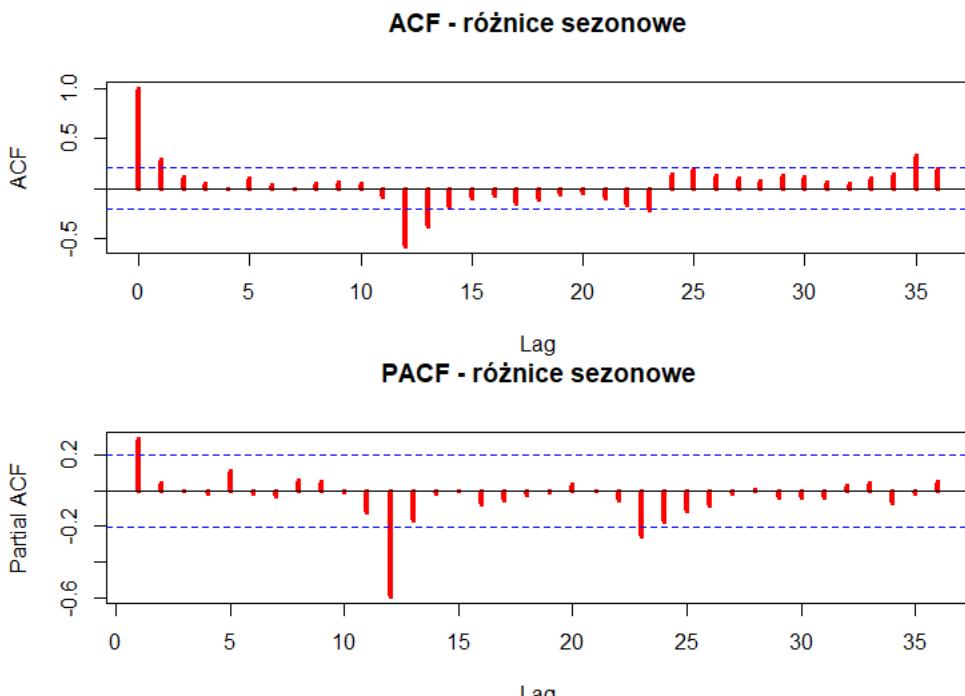
Źródło: Opracowanie własne w programie RStudio

Wykres 11. ACF i PACF dla pierwszych różnic regularnych cen ziemniaków w Polsce



Źródło: Opracowanie własne w programie RStudio

Wykres 12. ACF i PACF dla pierwszych różnic sezonowych cen ziemniaków w Polsce



Źródło: Opracowanie własne w programie RStudio

W celu formalnego zweryfikowania stacjonarności badanego szeregu przeprowadzony został test Dickeya-Haszy-Fullera (DHF), oraz zweryfikowane zostały wyniki testu Breusha-Godfreya, który pozwolił na przetestowanie założenia o braku autokorelacji. Przez wzgląd na miesięczny charakter badanego szeregu test ten musiał być spełniony do 6-tego rzędu włącznie. Początkowo przeprowadzana została analiza autokorelacji reszt dla zróżnicowanej sezonowo zmiennej, która wykazała brak istnienia autokorelacji reszt. W kolejnym kroku przeprowadzony został test Dickeya-Fullera, w celu potwierdzenia stacjonarności, dając p-value na poziomie 0.5, co może sugerować, że na poziomie istotności 5% istnieje brak podstaw do odrzucenia hipotezy zerowej, tym samym wyciągając wniosek o braku stacjonarności. Wartość ta jest jednak bardzo blisko przyjętego poziomu istotności, zatem dla uzyskania zapewnienia podjętej decyzji przeprowadzony został również test KPSS o odwrotnej hipotezie. Jego wyniki wykazały, że p-value przyjmuje wartości większe niż 0.1 a zatem istnieje brak podstaw do odrzucenia hipotezy zerowej o stacjonarności. Dodatkowo analiza wykresów ACF i PACF po różnicowaniu sezonowym wskazuje na brak wyraźnych oznak sezonowości i autokorelacji. Przez wzgląd na otrzymane wyniki zakładam zatem, że pojedyncze różnicowanie jest wystarczające do uzyskania stacjonarności, tym samym zarówno współczynnik d , jaki i D w modelu SARMIA będzie wynosił 1, przy częstotliwości s równej 12.

W kolejnym etapie sprawdzono, czy szereg pierwszych różnic sezonowych nie jest białym szumem. Do tej weryfikacji wykorzystane zostały testy Ljunga-Boxa oraz Boxa-Pierca, dające satysfakcjonujące wyniki. P-value dla obu z nich przyjęły dużo mniejsze wartości niż przyjęty 5% poziom istotności, czego potwierdzenie widoczne jest również na wykresie 12., ze względu na występujące wypustki, które są statystycznie istotne, a zatem dalsza analiza mogła zostać przeprowadzona.

W ramach identyfikacji modelu SARIMA określone zostały rzędy procesów AR (p) i MA (q) oraz rzędy procesów sezonowych SAR (P) i SMA(Q), ustalając postać SARIMY $(1,1,1)(1,1,1)_{12}$. Na jego podstawie przeprowadzona została procedura od ogólnego do szczególnego w celu znalezienia najlepszego, a tym samym najprostszego modelu w oparciu o kryteria AIC oraz BIC, której wyniki przedstawione są w tabeli 1. W pierwszym kroku przebadane zostały komponenty sezonowe w oparciu o model SARIMA $(0,1,0)(1,1,1)_{12}$. Analiza kombinacji modeli w oparciu o wyniki testu wiarygodności wykazała, że najlepszym modelem jest SARIMA $(0,1,0)(1,1,0)_{12}$. Wobec tego w kolejnym kroku w celu identyfikacji efektów regularnych postanowiono przeprowadzić na modelu SARIMA $(1,1,1)(1,1,0)_{12}$ ponownie metodę od ogółu do szczegółu. Analiza wykazała, że modelem z najniższym AIC oraz BIC jest SARIMA $(1,1,0)(1,1,0)_{12}$, dla której przeprowadzony został również testy Ljunga-Boxa oraz Boxa-Pierca, dla których p-value wynosiły odpowiednio 0.24 oraz 0.46, będąc tym samym większe od przyjętego 5% poziomu ufności. Zatem istnieje brak podstaw do odrzucenia hipotezy zerowej o tym że reszty są białym szumem.

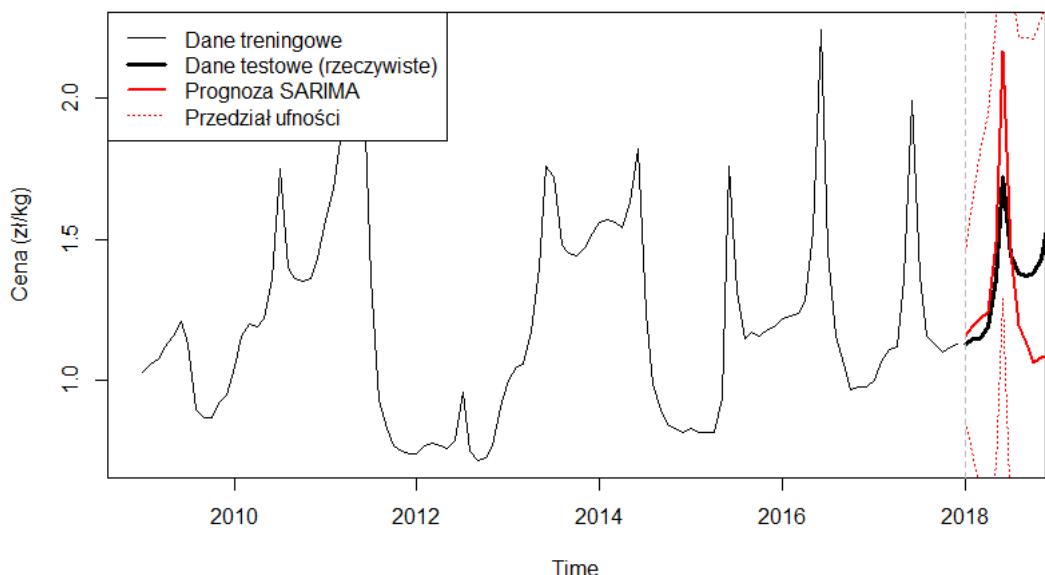
Tabela 1. Zestawienie wskaźników AIC i BIC dla estymowanych modeli klasy SARIMA

SARIMA	AIC	BIC
(1,1,1) (1,1,0)	-78.649	-68.434
(0,1,1) (1,1,0)	-79.671	-72.009
(1,1,0) (1,1,0)	-80.151	-74.490
(0,1,0) (1,1,0)	-78.835	-73.728

Źródło: Opracowanie własne

Wybór najlepszego modelu pozwala zatem przejść do prognozowania. Wykres 13. Przedstawia wartość przewidywań obserwacji wraz z jej realnymi wartościami. Czerwone przerywane linie przedstawią przedziały ufności, natomiast pogrubiona czerwona linia prezentuje prognozę uzyskaną za pomocą modelu SARIMA $(1,1,0)(1,1,0)_{12}$. Widać zatem, że przyjmuje ona podobny trend jak badany szereg, dodatkowo odwzorowując cykliczne zachowania obecne dla cen ziemniaków, natomiast amplituda wahań jest niedoszacowana oraz model nie przewiduje występowania gwałtownych skoków cenowych, co może być uzasadnione ponieważ okres ten lekko odbiega od poprzednich okresów. W kolejnym kroku obliczone zostały również błędy prognoz, które przedstawione zostaną w ogólnym zestawieniu dla wszystkich wykorzystanych metod.

Wykres 13. Prognozy cen ziemniaków w Polsce za pomocą modelu SARIMA (1,1,0) (1,1,0) dla okresu 2009-2018.



Źródło: Opracowanie własne w programie RStudio

Porównanie prognoz

W celu wyciągnięcia wniosków odnośnie wyboru najlepszego modelu dla analizy cen ziemniaków w Polsce w latach 2009-2018 stworzone zostało zestawienie błędów (MAE, RMSE i MAPE) dla metody naiwnej, średnich ruchomych, prostego wygładzania wykładniczego, modelu Holtta-Wintersa w wersji addytywnej oraz multiplikatywnej, modelu ETS oraz najlepszego modelu SARIMA. Zestawienie otrzymanych wyników przedstawione zostało w tabeli 2. Analiza wyników przedstawionych w tabeli wskazuje że najlepszy wynik osiągnęła metoda średnich ruchomych dla optymalnego okna równego 3 miesiące, dla której wartość RMSE wynosi 0. 0731, MAE wynosi 0. 0448 oraz MAPE wynosi 3.03%. Warto dodatkowo zauważyć, że to właśnie prostsze metody takie jak metoda naiwna oraz średnia ruchoma dużo lepiej poradziły sobie niż bardziej zaawansowane metody, dające zdecydowanie dużo gorsze dopasowanie. Najgorsze wyniki uzyskało SARIMA (1,1,0) (1,1,0) z RMSE równym 0.2568, MAE równym 0.1977 oraz MAPE równym 13.56%, co daje prawie 4-krotnie wyższe wyniki, pomimo teoretycznej przewagi wynikającej z uwzględnienia struktury autoregresyjnej i sezonowości. Może to natomiast wynikać bezpośrednio z tego jak różne metody traktują informację historyczną. Dla danych o wysokiej zmienności, w tym przypadku cena ziemniaka, która bardzo szybko odpowiada na wydarzenia rynkowe, informacja sprzed kilku miesięcy może nie być na tyle wartościowa, przez co większe znaczenie zaczynają odgrywać najnowsze trendy. Proste metody mogą zatem rzeczywiście stanowić silny benchmark dla bardziej zaawansowanych metod, czego dowodem są wyniki uzyskane w tym badaniu.

Tabela 2. Wartości RMSE, MAE i MAPE dla wybranych metod

METODA	RMSE	MAE	MAPE
Naiwna	0.1569	0.1054	0.0751
Średnia Ruchoma MA(2)	0.0784	0.0527	0.0375
Średnia Ruchoma MA(3)	0.0731	0.0448	0.0303
Średnia Ruchoma MA(6)	0.1244	0.0671	0.0439
Średnia Ruchoma MA(12)	0.1791	0.1355	0.0960
Proste wygładzanie wykładnicze	0.1502	0.0966	0.0644
Holt-Winters multiplikatywny	0.2617	0.2215	0.1628
Holt-Winters addytywny	0.2291	0.2153	0.1578
ETS (M,N,M)	0.2154	0.1752	0.1226
SARIMA (1,1,0) (1,1,0)	0.2568	0.1977	0.1356

Źródło: Opracowanie własne

W celu formalnego przetestowania która metoda wykazywała najlepszą trafność prognozy przeprowadzony został test Diebolda-Mariano. Wyniki przeprowadzonych 21 testów zawarte zostały w tabeli 3. Analiza wyników wykazała że dla 8 z 21 przypadków stwierdzono statystycznie istotne różnice na poziomie 5%. Dodatkowo wskazując na występującą hierarchię metod. Na jej szczycie znajduje się metoda średnich ruchomych dla optymalnego okna prognozy, wykazując dla wszystkich 6 metod istotną statystyczną przewagę. Drugą w hierarchii metodą jest proste wygładzanie wykładnicze, które w swojej prostocie wykazała przewagę nad bardziej zaawansowanymi metodami, jedynie w porównaniu do MA(3) okazało się ono gorsze. Trzecie miejsce zajęła natomiast metoda nawiązana. Natomiast metody takie jak: model Holta-Wintersa w wersji addytywnej i multiplikatywnej, ETS i SARIMA nie wykazały znaczących różnic między sobą. Wyniki te mogą być zaskakujące jednak w obliczu analizy cen ziemniaków w Polsce w latach 2009-2018, odznaczające się silną zmiennością oraz sezonowością, jak również prze krótki, 12-miesięczny horyzont prognozy wyniki te mogą być jak najbardziej adekwatne.

Tabela 3. Wyniki testu Diebolda-Mariano dla wybranych metod

Metoda 1	Metoda 2	p-value „twosided”	p-value „less”	p-value „greater”	Lepszy model
Naiwna	MA (3)	0. 0888	0.9555	0.0444	MA (3)
Naiwna	SES	0. 8615	0.5692	0.4307	Brak różnic
Naiwna	HW_add	0. 0140	0.0070	0.9929	Naiwna*
Naiwna	HW_mult	0. 0266	0.0134	0.9866	Naiwna *
Naiwna	ETS	0. 6609	0.3304	0.6695	Brak różnic
Naiwna	SARIMA	0. 2279	0.1139	0.8860	Brak różnic
MA (3)	SES	0.0739	0.0369	0.9630	MA (3)
MA (3)	HW_add	0.0000	0.0000	1.0000	MA (3) ***
MA (3)	HW_mult	0.0086	0.0043	0.9956	MA (3) **
MA (3)	ETS	0.0076	0.0038	0.9961	MA (3) **
MA (3)	SARIMA	0.0098	0.0049	0.9950	MA (3) **
SES	HW_add	0.0054	0.0027	0.9972	SES **
SES	HW_mult	0.0185	0.0092	0.9907	SES *
SES	ETS	0.2408	0.1204	0.8795	Brak różnic
SES	SARIMA	0.0581	0.0290	0.9709	SES
HW_add	HW_mult	0.3796	0.1898	0.8101	Brak różnic
HW_add	ETS	0.6978	0.6510	0.3489	Brak różnic
HW_add	SARIMA	0.4829	0.2414	0.7585	Brak różnic
HW_mult	ETS	0.4803	0.7598	0.2401	Brak różnic
HW_mult	SARIMA	0.9315	0.5342	0.4657	Brak różnic
ETS	SARIMA	0.0679	0.0339	0.9660	ETS

Źródło: Opracowanie własne

Zakończenie

Celem niniejszej pracy było przeanalizowanie sezonowego szeregu cen ziemniaków w Polsce w latach 2009-2018 przy użyciu różnych metod progностycznych oraz ich analiza porównawcza, pozwalająca na wyłonienie najefektywniejszego podejścia. W ramach badania zastosowane zostały metody takie jak: prognoza naiwna, średnia ruchoma, proste wygładzanie wykładnicze, model Holta-Wintersa w wersji addytywnej oraz multiplikatywnej, ETS oraz model SARIMA. Podejście to pozwoliło na uwzględnienie oraz porównanie szerokiego spektrum technik progностycznych, poczynając od tych prostych, przechodząc w dalszej części pracy do tych bardziej zaawansowanych. W początkowej części pracy przeprowadzona została dekompozycja szeregu przy użyciu metody Tramo-Seats, pozwalająca na dogłębną analizę struktury badanego szeregu. Dane podzielone zostały na próbę testową oraz estymacyjną, co pozwoliło na rzetelną ocenę jakości otrzymanych prognoz.

Po przeprowadzonym badaniu dla 7 technik przeprowadzona została analiza porównawcza na podstawie obliczonych błędów prognoz (RMSE, MAE i MAPE). W wyniku tego otrzymano dosyć zaskakujące wyniki, które sugerowały, że znaczącą przewagę w prognozowaniu cen ziemniaka mają proste metody progностyczne. Najlepszą metodą okazała się średnia ruchoma, dla optymalnego okna prognozy wynoszącego 3 miesiące, która odznaczyła się najmniejszymi błędami prognoz. Natomiast metodą z największymi błędami prognoz osiągającymi wartości powyżej 0.20 był model SARIMA, pomimo swojej przewagi pozwalającej na uwzględnienie sezonowości oraz struktury autoregresyjnej. Potwierdzenie wyciągniętych wniosków wykazała również analiza testu Diebolda-Mariano dla 21 przypadków. Test ten wykazał, że istnieją statystyczne różnice w jakości prognoz, jednocześnie sugerując na występującą przewagę średniej ruchomej dla optymalnego okna. Wyniki te w obliczu analizy cen ziemniaków w Polsce w latach 2009-2018, odznaczającymi się silną zmiennością oraz sezonowością, jak również przez wzgląd na krótki, 12-miesięczny horyzont prognozy mogą być uzasadnione. Warto również zwrócić uwagę na potencjalnie występującą niestabilność parametrów, takich jak: zmiany pogody, zmiany w technologiach uprawy, które powodują, że model SARIMA nie będzie odpowiednim wyborem. Porównując otrzymane wyniki z tymi przedstawionymi w artykule bazowym, w którym najlepszą metodą okazała się ARIMA pokonując model Holta-Wintersa widzimy że dla Polski sytuacja jest wręcz przeciwna, co może wynikać z występowania mniejszej zmienności na rynku tureckim.

W występujących warunkach niepewności producenci muszą jednak podejmować decyzje dotyczące planowanej wielkości uprawy, co jest nie tylko ryzykowne dla nich, ale także dla producentów oraz samych konsumentów. W tym kontekście wykorzystanie metody średniej ruchomej MA (3) może pomóc w otrzymaniu rzetelnych prognoz, oferując przy tym prostotę implementacji. Dzięki temu mogą one pomóc w podejmowaniu racjonalnych decyzji dla każdego z uczestników łańcucha dostaw. Dodatkowo przyczyniając się do kształtowania skutecznej polityki rolnej oraz systemów ostrzegania przed kryzysami żywieniowymi.

Pomimo potencjalnego wkładu w literaturę, przyszłe badania powinny rozszerzyć analizę o dodatkowe zmienne egzogeniczne, np. ceny paliw, warunki pogodowe, które z pewnością mają wpływ na cenę ziemniaków w Polsce. Uwzględnienie szerszej perspektywy pozwoliłoby na jeszcze lepsze opisanie zmienności badanego szeregu.

Powyższa analiza wykazała zatem, że proste metody mogą rzeczywiście być konkurencyjne dla bardziej zaawansowanych metod. Pomimo swojej prostoty to średnia ruchoma MA(3) okazała się być najefektywniejszym narzędziem do prognozowania cen ziemniaka w Polsce, podkreślając tym samym wartość metod klasycznych dla analizy rynków charakteryzujących się wysoką zmiennością.

Bibliografia

- Ministerstwo Rolnictwa i Rozwoju Wsi (2024). *Rynek ziemniaka*. Dostępne online: <https://www.gov.pl/web/rolnictwo/rynek-ziemniaka> (dostęp: 20.06.2025).
- Nowacki Wojciech (2024). *Historia ziemniaka*. Dostępne online: <https://polskiziemniak.pl/historia-ziemniaka/> (dostęp: 20.06.2025).
- Şahinli, M. A. (2020). Potato price forecasting with Holt-Winters and ARIMA methods: a case study. *American Journal of Potato Research*, 97(4), 336-346. DOI: 10.1007/s12230-020-09788-6